



Mechanistic Studies of Polycomb Group Proteins

Citation

Grau, Daniel James. 2012. Mechanistic Studies of Polycomb Group Proteins. Doctoral dissertation, Harvard University.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:10104005>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Mechanistic studies of Polycomb group proteins**Abstract**

Most cells within multicellular organisms contain the same genetic information, yet the appropriate tissue-specific expression of genes is required for the proper formation of adult tissues. Genes can either be “turned on” or “turned off” from the initial zygotic state and maintained during subsequent cell divisions. Maintaining the correct expression profiles during cell divisions is accomplished by a number of different nuclear factors. One of the key families of proteins that maintains the repression of target genes during development is the Polycomb group (PcG) of proteins. PcG proteins form a number of different multi-subunit protein complexes that interact with specific regions of chromatin and direct the repression of nearby genes by reducing transcription. One PcG complex, Polycomb repressive complex 1 (PRC1), inhibits transcription and nucleosome remodeling as well as compacts chromatin, both in vivo and in vitro. The in vitro repressive activities map mainly to one subunit of *Drosophila* PRC1—the Posterior sex combs (PSC) protein. The PRC1 complex is conserved in many other organisms including mammals. To better understand the mechanisms involved in PcG mediated repression we undertook a biochemical structure/function analysis of mouse PRC1. In chapter one, I review the current understanding of PcG biology and a rationale for the dissertation is provided. In chapter two, data are presented that argues that a mouse PRC1 protein, M33/Cbx2, which is non-homologous to PSC, is responsible for chromatin

compaction and repression of nucleosome remodeling. Data are presented that suggests these activities are localized to a basic, natively unfolded region of M33/Cbx2. In chapter three, we extend the findings from chapter two in an attempt to predict whether homologous PcG proteins from other species besides fly and mouse have biochemical activity. In agreement with predictions, a panel of recombinant PcG proteins was generated and data are presented that shows the predicted active PcG proteins are capable of both inhibition of nucleosome remodeling and compaction of chromatin. Finally, in chapter four, the implications of the data presented are discussed, and directions for further inquiry are explored.

Table of Contents

Abstract	iii
Table of Contents	v
List of Figures	viii
List of Tables	ix
Acknowledgements	x
Chapter One: Introduction	1
<i>Drosophila</i> Polycomb genetics	1
Mouse Polycomb genetics	7
Polycomb complexes	9
Histone post-translational modifications and Polycomb	14
PRC2 and histone methylation	15
Mechanisms of transcriptional repression by PcG proteins	18
Dissertation overview	22
Chapter Two: Identification of M33 as a functional homolog of	
Posterior sex combs	24
Preface	24
Introduction	25
Results	26
M33 is a functional homolog of PSC	26

None of the conserved features of M33 are required for in vitro activity	30
Positively charged residues are required for optimal activity in vitro	37
Compaction of chromatin by M33	46
Materials and Methods	49
 Chapter Three: Charge characteristics predicts activity of divergent PcG proteins	52
Preface	52
Introduction	53
Results	53
Evolution of PcG proteins suggests charge is a predictor for in vitro activity	53
Compaction of chromatin by divergent PcG proteins	62
Materials and Methods	65
 Chapter Four: Discussion and Conclusions	
Overview	66
Roles of natively unfolded proteins	66
Charged domains and PcG function	67
PcG protein evolution	69
Molecular nature of PcG/chromatin interactions	70
Technical challenges overcome	71
Future directions	73

Appendix A: 76

Grau DJ, Chapman BA, Garlick JD, Borowsky M, Francis NJ, Kingston RE. 2011. Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev* **25**: 2210-21.

Appendix B: 89

Grau DJ, Antao JM, Kingston RE. 2010. Functional dissection of Polycomb repressive complex 1 reveals the importance of a charged domain. *Cold Spring Harb Symp Quant Biol* 2010 **75**: 61-70.

References 100

List of Figures

Figure 1.1 *Homeotic* genes in *Drosophila* development

Figure 1.2 *Drosophila* and mammalian Polycomb complexes

Figure 2.1 Purification of *Drosophila* and mouse PcG proteins

Figure 2.2 M33 is the functional subunit of mouse PCC

Figure 2.3 Inhibition of remodeling requires nucleosomes and optimal salt concentrations

Figure 2.4 Purification of amino and carboxy-terminus truncation mutants of M33

Figure 2.5 Inhibition of remodeling by M33 truncation mutants

Figure 2.6 Inhibition of remodeling by M33 requires a basic region

Figure 2.7 Inhibition of remodeling activity of charge mutant PcG complexes and basic non-PcG proteins

Figure 2.8 Alignment of one of the regions of M33 where basic residues were mutated

Figure 2.9 M33 and intrinsic disorder.

Figure 2.10 Compaction of nucleosomal arrays by mouse PcG proteins

Figure 3.1 Analysis of evolutionary conservation of PcG function

Figure 3.2 Inhibition of remodeling activity of PcG proteins from other species

Figure 3.3 GST does not impact inhibition of remodeling activity

Figure 3.4 Compaction of nucleosomal arrays by diverse PcG proteins

Figure 4.1 A model for chromatin compaction by the mouse PRC1 core complex

List of tables

Table 1.1 Selected fly and mouse PcG proteins

Table 2.1 Summary of residue changes in M33 charge mutants

Table 3.1 Classification of predicted active and inactive PcG proteins

Acknowledgements

I wish to thank my advisor, Robert E. Kingston, for fostering an environment that is genuinely open and collegial, for being supportive through times both good and bad, for his generosity, and especially for being patient with graduate students with severe cases of writers block.

I wish to thank my dissertation advisory committee, Danesh Moazed, Anders Näär, and Ting Wu for always giving constructive advice during the advisory meetings that I did not schedule frequently enough.

I wish to thank members of the Kingston lab, past and present, for being a kind-hearted and friendly bunch. Anytime I had a question I felt like I could ask anyone without fear. I'm especially grateful for scientific discussions that took place in front of Simches 185.

I wish to thank the Molecular Biology Department and all of the employees who go above and beyond what is reasonably expected to be as helpful as possible.

I wish to thank the BBS administrators: Kate, Maria and Steve, for always having friendly faces, for always being helpful, and for never being snarky when graduate students do not schedule DAC meetings on time.

I would like to thank my family for always supporting me in whatever I have pursued.

And finally, I would like to thank Amy Ma for her strength, for her tenderness, for her devotion, for her patience, for her generosity, and most importantly for her impeccable taste.

Chapter One: Introduction

In higher eukaryotes, gene expression profiles must be established early in developing embryos and then be faithfully transmitted from progenitor cells through multiple cell divisions. The appropriate gene products are required to establish a particular tissue type; this is accomplished by either up-regulation or down-regulation of gene products. Proteins that up-regulate transcription are called activators, whereas proteins that down-regulate transcription are called repressors. Two key classes of protein complexes that accomplish this during development are the Polycomb group (PcG) and the Trithorax group (trxG). TrxG proteins are involved in gene activation whereas PcG proteins are involved in gene repression. One key function of trxG and PcG proteins is the maintenance of homeotic (*Hox*) gene expression patterns. It is now clear that trxG and PcG protein complexes function mainly by interacting with chromatin. Chromatin, at the most basic level, is the association of DNA and a core octamer of histone proteins. One way the cell modulates gene expression is by modifying the organization of chromatin to either promote or repress transcription. TrxG proteins have several enzymatic activities that modify chromatin and lead to gene activation whereas PcG proteins have both enzymatic and non-enzymatic activities that modify chromatin and lead to transcriptional repression. The rest of this introduction will focus on the biology of PcG protein function and the rationale for the experiments that are presented in chapters two and three.

***Drosophila* Polycomb genetics**

For over one hundred years *Drosophila melanogaster* has been used as a model organism for the study of biology. Given the speed at which flies reach adulthood, the ease of husbandry, relatively complicated body-plan, and evolutionarily conserved genes,

Drosophila makes an ideal model system for the study of development. Early development in *Drosophila* begins with anterior-posterior gradients of mRNAs and proteins that are maternally deposited (Figure 1.1A) (Driever and Nusslein-Volhard 1988a; Driever and Nusslein-Volhard 1988b). These gradients establish the expression patterns of *gap* and *paired-rule* genes, collectively known as the segmentation genes. The segmentation gene products initiate the anterior-posterior expression patterns of the homeotic (*Hox*) genes in segments; these patterns are subsequently maintained throughout development (Akam 1987). In *Drosophila*, the *Hox* genes are localized in two different clusters: The Bithorax Complex (BX-C), a large (~315 kb) region that encodes three homeotic genes, and the Antennapedia Complex (ANT-C), which encodes five genes (Figure 1.1B) (Lewis 1978; Duncan 1987; Kaufman et al. 1990). The homeotic genes encode DNA-binding transcription factors that regulate the anterior-posterior development of *Drosophila* by activating target genes in only the proper segments of developing embryos. Misexpression of a *Hox* gene in segments where it is normally inactive leads to transformations of that segment to a fate that is more anterior or posterior. For example, mutations in one *Hox* gene, *Sex combs reduced* (*Scr*), lead to the loss of bristles (sex combs) that are normally found only on the front legs of adult male flies (Kaufman et al. 1990; Pattatucci et al. 1991). Ectopic expression of *Scr* leads to the formation of sex combs on the second and third sets of legs on male flies, a location where they are normally absent (Pattatucci and Kaufman 1991; Pattatucci et al. 1991).

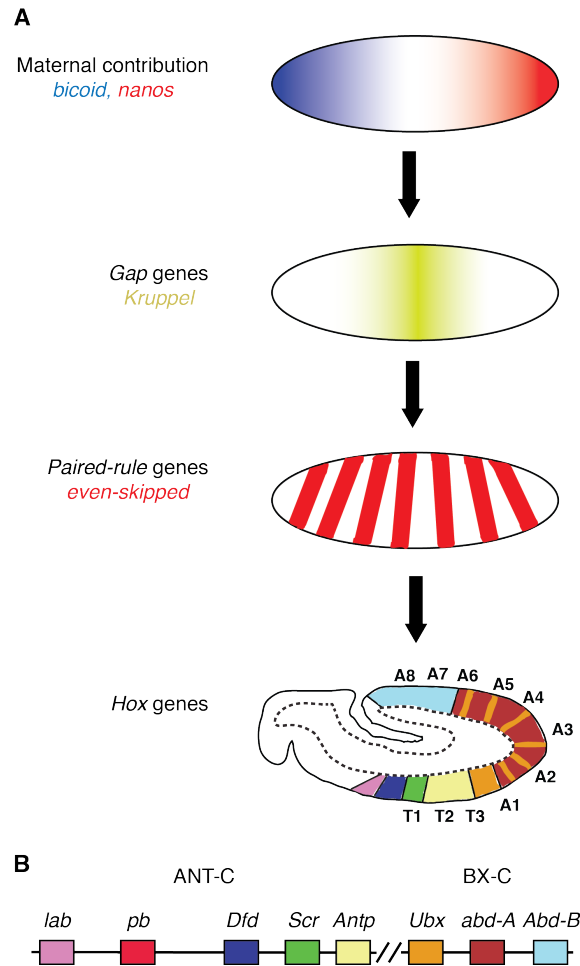


Figure 1.1 *Homeotic* genes in *Drosophila* development

A. Examples of patterning formation during *drosophila* embryonic development. Initial gradients are formed using maternally contributed mRNAs and proteins. These gradients initiate the patterns of the *gap* and *paired-rule* genes, which help to set up the initial boundaries of expression of the *Homeotic* (*Hox*) genes. *Hox* gene expression boundaries are subsequently maintained by the Polycomb and trithorax group proteins. B. Genomic organization of the two *Drosophila* *Hox* clusters: Antennapedia Complex (ANT-C), and Bithorax Complex (BX-C). Colors correspond to boundaries of expression as depicted in panel A. Based on Figure 1 of *Generating patterns from fields of cells* (Sanson 2001).

The segmentation genes are only active early in development, yet the *Hox* genes expression patterns are maintained through adulthood. How might the expression patterns be maintained? In 1947, P. Lewis isolated a mutant fly with ectopic sex combs on the second and third sets of legs and named this mutant *Polycomb* (*Pc*) (Lewis 1947). Homozygous mutants of *Pc* are embryonic lethal. Later on it was noted by E.B. Lewis that the homozygous mutant embryos had segments that resembled those that were more posterior and hypothesized that the function of *Pc* is to repress *Hox* genes in trans (Lewis 1978).

Since the segmentation genes establish the initial patterns of *Hox* genes, and PcG genes are required for *Hox* gene silencing throughout the rest of development, the Polycomb group is considered to be involved in the maintenance, rather than establishment, of gene repression (Struhl and Akam 1985; Simon et al. 1992). Possible mechanisms for how *Pc* mediated gene silencing is maintained during cell division will be discussed below.

Since the initial isolation of the *Pc* mutant, numerous other mutants have been isolated with the extra sex combs phenotype (Muller and Verrijzer 2009). Genes with mutants that have either the extra sex combs phenotype or enhance the phenotype of *Pc* alleles make up the Polycomb group (PcG) (Jürgens 1985). Based on the number of deficiencies that do not complement PcG mutants, it has been estimated that there are approximately 40 PcG genes (Jürgens 1985). Currently, 27 different complementation groups have been isolated, and the genes that have been molecularly characterized encode for proteins that form multi-subunit complexes involved in gene silencing (Gaytan de Ayala Alonso et al. 2007).

Like *Pc*, homozygous loss-of-function mutations in the other PcG genes are embryonic lethal with posterior transformations of segment identity (Lewis 1978). Trans-heterozygotes of two or more PcG mutants are more severe than either of the heterozygotes alone, suggesting that the gene products function together (Sato et al. 1983). The rest of this section will focus on the genetics of *Pc*, *Posterior sex combs* (*Psc*), and the corresponding genes that have been characterized in knockout mice (For sake of brevity I will only discuss *Psc* and not the functionally related paralog *Su(z)2*).

Since the isolation of the original *Pc* mutant in 1947, at least 50 additional alleles have been described, of which ten have been molecularly characterized (Messmer et al. 1992). These alleles all have the *Pc* mutant phenotype: embryonic lethal when homozygous and ectopic sex combs on the second and third legs when heterozygous. The PC protein is 390 amino acids and contains two conserved domains: a N-terminal chromodomain and a C-terminal C-box (Paro and Hogness 1991). The chromodomain is common domain found in chromatin interacting proteins and binds to methylated histone tails (Lachner et al. 2001). The C-box is a protein-protein interaction motif that is required for interactions with other PcG proteins that contain RING domains (Wang et al. 2008). Most of the characterized lesions that disrupt *Pc* function occur in either the chromodomain or C-box (Franke et al. 1995). In immunohistochemistry experiments, mutations in the chromodomain disrupt normal binding to chromatin, whereas the C-box is dispensable for localization (Messmer et al. 1992).

The first characterized *Posterior sex combs* mutant, *Psc^l*, is embryonic lethal when homozygous and has head defects, as well as transformations of the head and thorax to abdominal structures (Jürgens 1985). In heterozygotes, extra sex combs are

sometimes observed. *Psc* encodes for a 1603 amino acid protein and also contains two conserved domains near the N-terminus of the protein: a C3HC4 zinc finger RING domain, and a helix-turn-helix-turn-helix (HTH) motif (Brunk et al. 1991). As with *Pc*, a number of alleles have been characterized molecularly. In one study, an allelic series of *Psc* mutants that has both moderate (hopeful) and strong (hapless) phenotypes was sequenced and characterized biochemically (Wu and Howe 1995; King et al. 2005). Hapless alleles are those that are lethal when in trans to other hapless alleles, whereas hopeful alleles are viable in trans to other hopeful alleles.

Most of the mutant alleles sequenced have stop codons introduced within the coding sequence that result in truncations of PSC from the C-terminus (King et al. 2005). In general, there is a good correlation between the extent of truncation and the severity of the phenotype. Additionally, the activity of mutant PSC proteins in biochemical assays correlates well with the severity of phenotype. Specifically, the larger C-terminal truncations are less active in assays that measure the ability of PcG proteins to inhibit remodeling and transcription in vitro—two activities that are thought to be biologically relevant for PcG function in vivo. This study highlights the importance of the extended C-terminus of PSC for both in vivo and in vitro function. However, in addition to the truncation mutants, there is a point mutant, *Psc^{e23}* (C268Y), that does not behave in accordance to the rest of the allelic series. The *Psc^{e23}* mutant encodes for a full-length protein with a single amino acid change in the second cysteine of the C3HC4 RING finger domain. This mutant is interesting because although it falls into the hapless group and is homozygous lethal, the recombinant protein is fully active in vitro. This suggests

that the C-terminal region of PSC is required for both in vitro and in vivo function, whereas the RING domain is required only for in vivo viability.

Mouse Polycomb genetics

In 1991 the *Bmi-1* gene was identified as a proto-oncogene in transgenic mice over-expressing the *myc* gene (Haupt et al. 1991; van Lohuizen et al. 1991b). In these animals over-expression of *myc* eventually leads to tumors, however, infection with Molony murine leukemia virus greatly accelerates the formation of tumors. The integration locations of the virus were mapped and nearly half of the integrations were found near the *Bmi-1* locus, leading to increased expression of *Bmi-1*. Later that same year the predicted protein sequences of *Psc* and *Su(z)2* were reported and it was established that *Bmi-1* is a homolog of these PcG genes (Brunk et al. 1991; van Lohuizen et al. 1991a). Since then, homologs for many of the *Drosophila* PcG genes have been isolated from mammalian tissues, and with current genomic information it is clear that mammals have homologs for most of the described fly genes. In contrast to flies, mammals generally have multiple paralogs of each fly PcG gene; I will briefly describe the phenotypes of the mouse knockouts for the homologs of *Psc* and *Pc*.

Bmi-1 was the first mouse PcG that was targeted for deletion (van der Lugt et al. 1994). *Bmi-1* null mutant mice survive through embryogenesis and die shortly after birth. *Bmi-1*^{-/-} mice exhibit mild posterior transformations of the axial skeleton that correspond to anterior shifts of *Hox* gene expression. Additionally, *Bmi-1*^{-/-} mice have neurological abnormalities that result in partial paralysis as well as defects in haematopoiesis and proliferation. Knockout of another *Psc* and *Su(z)2* homolog, *Mel-18*, leads to similar transformations of the axial skeleton, and similar defects in

haematopoiesis (Akasaka et al. 1996). However, there are differences: *Mel-18* null mice have intestinal defects. *Bmi-1*^{-/-};*Mel-18*^{-/-} double knockouts exhibit a much more severe phenotype and die as early embryos, whereas mice null for one gene and heterozygous for the other have intermediate effects (Akasaka et al. 2001). These results suggest that there exists some redundancy between the two proteins. However, while *Bmi-1* is an oncogene, *Mel-18* has been reported to have tumor suppressor activity (Kanno et al. 1995). This, along with the differences in the knockouts, suggests that although there are overlapping functions, there exists specialization for the functions of each protein. There are four additional PSC homologs in mice; knockouts have not yet been described.

Mice have five homologs of fly *Pc*, of which two have been knocked out (Core et al. 1997; Forzati et al. 2012). *M33/Cbx2* was the first mouse *Pc* homolog to be cloned (Pearce et al. 1992; Core et al. 1997; Forzati et al. 2012). *M33* null knockout mice develop to term and within a few hours of birth half of the animals die, and within 4 weeks post birth 90% of the animals die with growth defects. Like *Bmi-1* and *Mel-18* mutants, *M33*^{-/-} mice have defects in the axial skeleton due to derepression of *Hox* genes. In addition, the mutants have thymus and spleen defects with a pronounced decrease in the number of T cells. Cultured fibroblasts and splenocytes have proliferation defects. *M33*^{-/-};*Bmi-1*^{-/-} double mutants enhance the posterior transformations of the axial skeleton due to a greater derepression of *Hox* genes (Bel et al. 1998). Another phenotype that was not initially reported is a male-to-female sex reversal in null mice (Katoh-Fukui et al. 1998). Genetically male, *M33* null mice fail to form external genitalia and instead 50% develop uteri, oviducts and ovaries. Of the rest of the animals, 25% had both testis and ovaries. The significance of this phenotype is manifested by a clinical case in which

an XY patient presented with male-to-female sex reversal (Biaison-Lauber et al. 2009). Analysis of the patients *Cbx2* gene indicated two polymorphisms that result in P98L and R443P substitutions. A molecular explanation for both the mouse and human sex-reversal phenotype may be the misexpression of the Y-chromosome gene *Sry*. In *M33*^{-/-} mutant gonads *Sry* and *Sox9* expression is affected, and the sex-reversal phenotype is rescued by forced expression of *Sry* and *Sox9* (Katoh-Fukui et al. 2011).

Collectively, the phenotypes of the mouse PcG knockouts demonstrate that the role of PcG proteins in *Hox* gene regulation has been conserved. The increased severity of phenotypes in double mutants demonstrates that dosage effects with PcG proteins occurs both in flies and in mammals. In addition to common phenotypes, each of the PcG knockouts develops unique abnormalities, suggesting that there is regulation of more than just the *Hox* clusters. Indeed, genome wide mapping of PcG binding in flies and mice demonstrates that PcG proteins are involved in the regulation of hundreds of genes in addition to those located at the Hox clusters (Boyer et al. 2006; Bracken et al. 2006; Schwartz et al. 2006).

Polycomb complexes

Based on observed genetic interactions in flies it was proposed that PcG proteins may exist in multimeric protein complexes (Jürgens 1985). Indeed, currently there are five PcG complexes that have been identified in *Drosophila* (Beisel and Paro 2011). In general, these protein complexes contain both enzymatic and architectural factors. The five isolated PcG complexes are: dRing-associated factors (dRAF), Polycomb repressive deubiquitinase (PR-DUB), Pho repressive complex (PhoRC), Polycomb repressive complex 2 (PRC2), and PRC1 (See table 1.1 and Figure 1.2 for a comparison of these

complexes in flies and mammals) (Shao et al. 1999; Kuzmichev et al. 2002; Muller et al. 2002; Klymenko et al. 2006; Lagarou et al. 2008; Scheuermann et al. 2010). Of these, homologous complexes have been characterized in mammals for all of them except for PhoRC, although homologous proteins exist.

What is the protein composition and activities of these PcG complexes? The dRAF complex contains two canonical PcG proteins: dRING (also known as Sex combs extra (Sce)) and Posterior sex combs (PSC), as well as dRAF1/dKDM2, dRAF2, and mTor (Lagarou et al. 2008). dRING is an E3 ubiquitin ligase that stimulates the monoubiquitination of histone H2A at lysine 118 (H2AK118ub1), a mark that is associated with silent chromatin domains. dKDM2 is a lysine demethylase that removes the activating H3K36me2 mark. Additionally, dKDM2 seems to stimulate the ubiquitin ligase activity of dRING, thereby linking the removal of an activating mark with the deposition of a repressive mark.

PR-DUB is a recently characterized complex that contains the proteins Additional sex combs (Asx) and Calypso, and catalyzes the removal of H2AK118ub1 (Scheuermann et al. 2010). This mark is associated with regions of the genome that are normally repressed (Wang et al. 2004; Cao et al. 2005). Paradoxically these two PcG complexes have opposing enzymatic functions, the implications will be discussed below.

PhoRC contains the sequence-specific DNA-binding protein pleiohomeotic (PHO) and *Drosophila* Scm-related gene containing four mbt domains (dSfmbt)

Table 1.1 Selected fly and mouse PcG proteins

<i>Drosophila</i>	<i>Mouse</i>	Human	Complex	Domains	Functional role(s)
Enhancer of zeste (E(z))	Ezh1 Ezh2 (KMT6, Enx-1)	EZH1 (KMT6B) EZH2 (KMT6, ENX1)	PRC2 PRC2	SET SET	H3K27 Mase
Suppressor of zeste 12 (Su(z)12)	Suz12	SUZ12 (CHET9, JUAZ1)	PRC2		Enhances Mase activity of E(z)
Extra sex combs (Esc)	Eed (lusk)	EED (WAIT1)	PRC2	WD40 repeats	Enhances Mase activity of E(z)
Nucleosome remodeling factor 55 (Nurf55/Caf-1) Polycomb-like (Pc)	Rbpb4 (Caf-1, mRbAp48) Phr1 (Pcl1) Mtf2 (Pcl2) Phr19 (Pcl3) Cbx2 (M33, MOD2, Pc1) Cbx4 (Pc2) Cbx6 Cbx7 Cbx8 (Pc3)	RBBP4 (CAF-1, NURF55, RBAP48) PHF1 (PCL1) MTF2 (PCL2) PHF19 (PCL3) CBX2 (PC1) CBX4 (PC2) CBX6 CBX7 CBX8 (PC3)	PRC2 Interacts with PRC2	WD40 repeats PHD fingers and Tudor domain	Nucleosome binding Recruitment, enhances Mase activity of PRC2
Polycomb (Pc)	Phc1 (Edr, Mph1, Rae-28) Phc2 (Edr2, Mph2, p36) Phc3 (Edr3, Hph3)	PHC1 (EDR1, HPH1, RAE28) PHC2 (EDR2, HPH2, PH2) PHC3 (EDR3, HPH3)	PRC1 PRC1 PRC1	Chromodomain, AT-Hook and Pc-box Chromodomain and Pc-box Chromodomain and Pc-box Chromodomain and Pc-box	Binds H3K27triMe Binds H3K27triMe Binds H3K27triMe Binds H3K27triMe
Posterior sex combs (Psc) and Suppressor of zeste 2 (Su(z)2)	Bmi1 (Pcgl4) Mel18 (Pcgl2, Rnf110) Pcgl1 (Nspc1)	BMI1 (PCGF4, RNF51) MEL18 (PCGF2, RNF110, ZNF144) NSPC1 (PCGF1)	PRC1 PRC1 BCOR	RING finger RING finger RING finger	Enhances Ubiquitination Compacts nucleosomes
Polyhomeotic proximal (Ph-p) and distal (Ph-d)	Ring1 (Ring1A) Rnf2 (Ring1B, dinG)	RING1 (RING1A, RNF1) RNF2 (DING, RING1B)	PRC1/BCOR PRC1/BCOR	SAM SAM SAM	Spreading? Spreading? Spreading?
Sex combs extra (Sce/dRing)	YY1 (NF-E1)	YY1 (DELTA, INO80S, NF-E1)	PhoRC	RING finger RING finger	E3 ligase E3 ligase
Pleiohomeotic (Pho)	mSimbt (Smr)	SFMBT1 (RU1)	PhoRC	Zn fingers	Recruitment?
dSimbt	Bap1	BAP1	PR-DUB	Zn finger, MBT and SAM	Binds methylated histones
Calypso	Asx1	ASXL1	PR-DUB	UCH	Deubiquitinates H2A
Additional Sex Combs (Asx)	Bcor	BCOR	BCOR	Ankyrin repeat	Enhances Calypso activity
dKDM2	Kdm2b (Fbx10)	KDM2B (JHDM1B, FBXL10, CXXC2)	BCOR	CXXC, PHD finger, F-box and Leucine rich repeats	Targeting H2A ubiquitination?
dRYBP	Rybp (DEDAF, YEAF1)	RYBP (DEDAF, YEAF1)	BCOR	Zn finger	H3K36 demethylase
skpA	Skp1a (EMC19, OCP2, SKP1A, p19A)	SKP1 (EMC19, OCP2, SKP1A, p19A)	BCOR	SKP1	Protein-protein interaction

^a Paranthases denotes other protein names

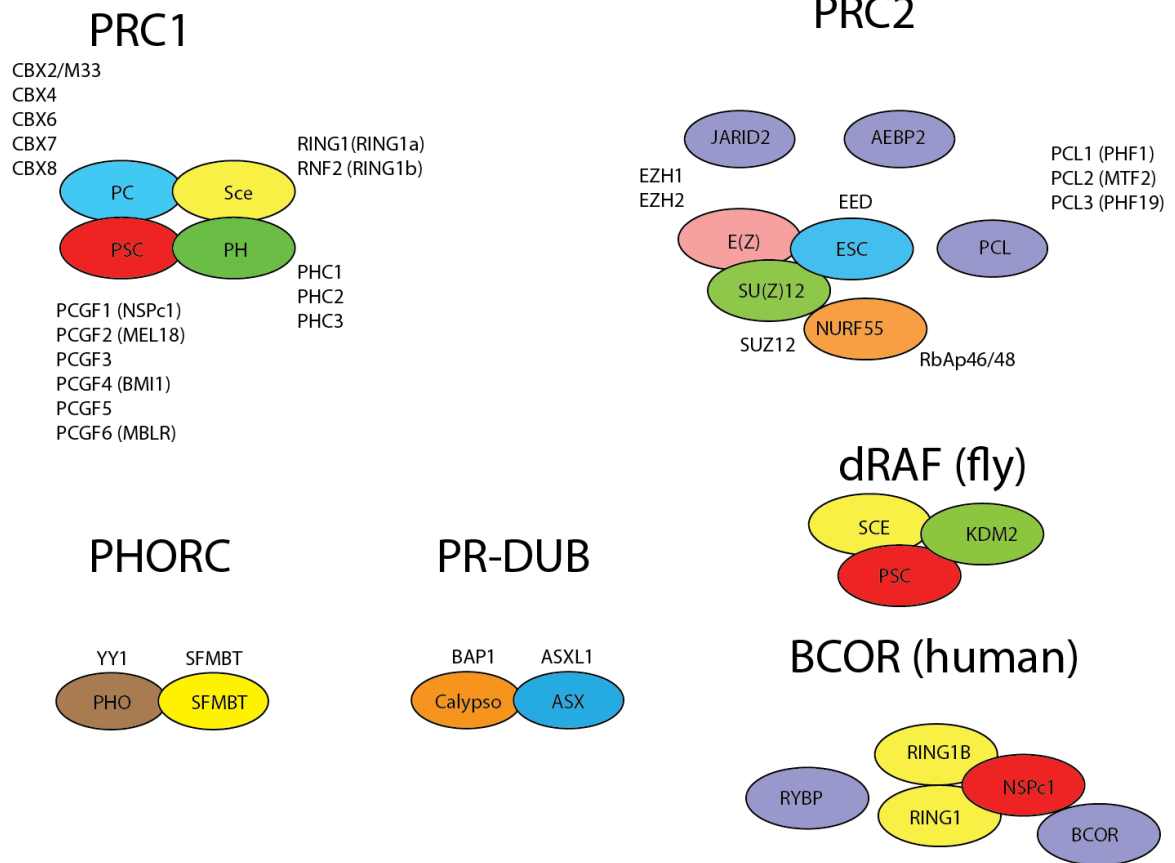


Figure 1.2 *Drosophila* and mammalian core Polycomb complexes

Ovals depict the subunits of the various *Drosophila* Polycomb core complexes except where specified. Text next to the ovals indicates the homologous mammalian Polycomb proteins. Ovals shaded in purple represent factors that are ancillary to the core complexes.

(Klymenko et al. 2006). Pho mediates the recruitment of the complex to Polycomb repressive elements (PREs) and dSfmbt binds to methylated histone H3 and leads to repression of *HOX* genes. PRC1 and PRC2 were the first PcG complexes to be isolated. In *Drosophila*, the core of PRC2 contains the subunits Extra sex combs (ESC), Enhancer of zeste (E(z)), Suppressor of zeste 12 (Su(z)12) and Nucleosome remodeling factor 55 (NURF55) (Kuzmichev et al. 2002; Muller et al. 2002). The SET domain of E(z) catalyzes the di- and tri-methylation of histone H3 lysine 27 (H3K27me₃), a repressive mark that will be discussed below. In mammals there also exists PRC2 complexes containing homologs of the core subunits found in *Drosophila*. Additionally, a number of accessory proteins have been found to interact with mammalian PRC2. Other subunits that are found associated with the mammalian complex are Phd-finger protein 1 (PHF1) and JARID2 (Sarma et al. 2008; Peng et al. 2009; Shen et al. 2009; Pasini et al. 2010). PHF1 has been implicated in enhancing the methyltransferase activity of PRC2. JARID2 is a protein that is currently under a lot of scrutiny. JARID2 contains a catalytically dead demethylase, owing to mutations in the catalytic domain, and is required for recruitment of PRC2 to chromatin targets (Pasini et al. 2010). Currently it is unclear whether or not JARID2 enhances the methyltransferase activity of PRC2, however, it remains a protein of much interest (Peng et al. 2009; Shen et al. 2009).

Structural work has implicated H3K27me₃ as a binding substrate for the chromodomain of Polycomb (PC), a component of PRC1 (Fischle et al. 2003; Min et al. 2003). PRC1 is a megadalton protein complex that is generally thought of as the PcG complex that is the driving force behind transcriptional repression. A functional core complex of PRC1, called Polycomb core complex (PCC), contains the proteins PC,

Polyhomeotic (PH), Posterior sex combs (PSC) and Sex combs extra (Sce/dRing) (Francis et al. 2001). Several non-enzymatic functions of PCC have been demonstrated in vitro. PCC can inhibit nucleosome remodeling by the SWI/SNF complex on reconstituted nucleosomal arrays (Francis et al. 2001). Additionally, PCC can inhibit transcription off of similar templates (King et al. 2002). Incubation of PCC with nucleosomal arrays leads to compaction of the nucleosomes from the canonical “beads on a string” morphology to more compact structures (Francis et al. 2004). Interestingly, the PSC subunit of PCC is able to confer all of these biochemical activities on its own. As mentioned above, recombinant PSC protein corresponding to genetically isolated alleles has in vitro activity that correlates with the phenotypic data, suggesting that the in vitro data is predictive of PSC function in vivo (King et al. 2005).

Mammalian PRC1 has also been purified from HeLa extracts, but less is known about the biochemical functions of mammalian PRC1 (Levine et al. 2002). A core complex from mouse, mPCC, inhibits remodeling in vitro (Lavigne et al. 2004).

Histone post-translational modifications and Polycomb

Within the cell are a host of different enzymes that can covalently modify proteins. These modifications can have profound effects on the localization or activities of target proteins. Within the nucleus are enzymes that specifically modify nucleosomes to regulate transcriptional profiles (Suganuma and Workman 2011). The nucleosome exists as an octamer of core histones that is wrapped with 147 base pairs of DNA. The canonical nucleosome is composed of two dimers of histones H2A and H2B, and a tetramer of histones H3 and H4. Much of the histone octamer exists as a globular core that is surrounded by the aforementioned DNA, however, the N-terminal ‘tails’ of the

histones are outside of the core--as well as C-terminal regions of H2A and H2B--and are unstructured in solution. The histone tails and certain locations within the octamer core can be covalently modified by enzymes that acetylate, methylate, phosphorylate or ubiquitinate at specific amino acid residues. Additionally, enzymes that remove each these modifications have been characterized. It is well established that certain modifications correlate well with the transcriptional status of nearby genes. These modifications can affect transcriptional activity either by directly affecting chromatin structure and/or by recruitment of proteins. The precise mechanisms by which most of the various histone modifications direct transcriptional regulation remain unclear. Polycomb and Polycomb-like protein complexes can catalyze both the deposition or removal of histone tail modifications as well as interact specifically with modifications.

PRC2 and histone methylation

In the mammalian PRC2 complex, EZH2, and to a lesser extent the paralogous protein EZH1, catalyze the di- and tri-methylation of histone H3 lysine 27 (H3K27me_{2/3}) (Kuzmichev et al. 2002; Muller et al. 2002; Shen et al. 2008). This modification is associated with regions of the genome where transcription is repressed. Additionally, a non-enzymatic function of PRC2 may be to help maintain repressive chromatin domains by directly interacting with and compacting chromatin. Chromatin compaction is observed in vitro when EZH2 and EZH1 are incubated with chromatin substrates (Margueron et al. 2008). Recently it was demonstrated that PRC2 binds to the H3K27me₃ mark and this binding helps to enhance the methyltransferase activity of PRC2, likely through an allosteric mechanism (Hansen et al. 2008). The authors propose

a model where PRC2 binding to H3K27me3 acts a positive feedback loop to maintain methylation patterns after cell division.

It has been hypothesized that the enzymatic activities of PcG proteins could potentially target non-histone substrates. Recently, He et al. demonstrated that PRC2 can specifically methylate the GATA4 transcription factor (He et al. 2012). This methylation reduces the protein activity by preventing activation by p300. It will be interesting to determine what other non-histone substrates are targeted by PRC2 and other PcG enzymes.

The function of EZH1 is currently under debate. EZH1 is found in a complex with similar subunits as PRC2 and has high homology to EZH2, suggesting that it may have redundant functions (Margueron et al. 2008; Shen et al. 2008). Consistent with this, EZH1 can compact chromatin in vitro and can inhibit transcription when recruited to a reporter. However, a recent report using ChIP-Seq against EZH1 in undifferentiated and myogenic cell populations demonstrated that in differentiated populations EZH1 was associated with H3K4me3, PolII phosphorylated at serine 2, and active transcription (Mousavi et al. 2011). Knockdown of EZH1 led to the reduction of transcription at target genes, while overexpression increased the transcription of target genes. These differences in experimental outcomes will require further experimentation to address.

Methylation of H3K27 has been proposed to contribute to the recruitment of PRC1 through direct interactions of the modified histone tail with chromodomain (Cbx) proteins in PRC1 (Fischle et al. 2003; Min et al. 2003). However, it remains unclear the extent that H3K27me3 directs recruitment, as not all PRC1-bound locations are associated with the H3K27me3 modification (Schwartz et al. 2006). This issue is

confounded by the fact that in mammals there exists five PcG Cbx proteins (Cbx2, 4, 6, 7 and 8), and multiple paralogs of the other core subunits. Recently, two papers investigated the role of Cbx proteins in mouse stem cells during differentiation and came to similar conclusions. They found that in undifferentiated stem cells, the Cbx7 paralog is predominantly expressed and maintains pluripotency by repressing differentiation genes within the context of PRC1 (Morey et al. 2012; O'Loghlen et al. 2012). Cbx7 also mediates the repression of Cbx2, 4 and 8. Upon differentiation, the other Cbx proteins are up regulated, incorporated into PRC1 complexes, and inhibit both pluripotency genes and Cbx7. Therefore, multiple Cbx paralogs function in antagonistic roles at different times during differentiation. Detailed analysis of the paralogs of the other components of PRC1 will be necessary to determine whether they use similar mechanisms.

The role of ubiquitination is not entirely clear. Some genomic targets of PcG complexes containing Ring1B do not require the E3 ligase activity of Ring1B (Eskeland et al. 2010). Loss of Ring1B results in decompaction of *Hoxb* and *d*, and is rescued by catalytically inactive Ring1B. While the H2AK119ub1 mark is generally associated with silent chromatin, it may not be required at all loci. Recently another PcG complex has been described that deubiquitinates H2A (PR-DUB) (Scheuermann et al. 2010). In the absence of the deubiquitinase, PcG targets are derepressed. In compound mutants for both the H2AK118 ubiquitinase and deubiquitinase (Sce and Calypso) the mutant phenotype is enhanced over either mutant alone (Gutierrez et al. 2012). Paradoxically the repression of target genes requires a balance of both ubiquitinase and deubiquitinase activity. The authors speculate that while mono-ubiquitination of H2A by Sce is required for repression of target genes, the deubiquitinase may be required either to keep the

global pool of ubiquitin at a level that is usable by Sce, or to remove ubiquitinated H2A from PcG targets that repress by other mechanisms.

Mechanisms of transcriptional repression by PcG proteins

The precise mechanistic details of PcG mediated repression remain undetermined. There are several issues to consider regarding the mechanisms of PcG repression: 1) How are PcG complexes recruited to target loci? 2) What do the PcG complexes do when they are at target loci? 3) How is the repression program transmitted through many cell divisions? I will attempt to address the current understanding of these currently unresolved questions.

The first issue to consider is how PcG complexes are recruited to the appropriate targets during development. It has long been known that there are DNA elements proximal to the *Hox* genes that act as sites of PcG protein nucleation. These sites, called Polycomb repressive elements (PREs), are generally composed of several different elements of around a hundred base pairs, and are necessary and sufficient for PcG mediated silencing (Simon et al. 1993). It has been suggested that PcG proteins are recruited to PREs by the combination of multiple low-affinity interactions with DNA-binding proteins. In support of this, PREs contain many binding sites for specific DNA-binding proteins. The best-characterized binding sites in PREs are for Pleiohomeotic (PHO), GAGA factor (GAGA), and ZESTE (Muller and Kassis 2006). PHO, the *Drosophila* homolog of mammalian Yin Yang-1 (YY1), was the first PcG protein with specific DNA-binding to be characterized (Shi et al. 1991; Brown et al. 1998). PHO binds to hundreds of different sites on polytene chromosomes and overlaps with PSC at about 65% of the locations (Brown et al. 2003). PHO can interact with and recruit a

PRC1 complex in vitro, likewise YY1 has been found associated with both PRC1 and PRC2 (Garcia et al. 1999; Satijn et al. 2001; Mohd-Sarip et al. 2005). Despite considerable effort, it remains unclear how the DNA-binding proteins may recruit PcG proteins to PREs. Few specific DNA-binding proteins have been found stably associated in PcG complexes. Whether this is due to low-affinity contacts that do not survive purification is unknown. Additionally, despite PcG binding to hundreds of loci in the mammalian genome, only one mammalian PRE has been described (Woo et al. 2010). Whether mammals use different methods of recruitment is unknown.

The second issue to consider is the mechanisms PRC1 uses to repress transcription after it is recruited. Several different mechanisms of transcriptional repression by PcG proteins have been proposed: inhibiting the transcriptional machinery by preventing the binding of factors (steric hindrance), directly interacting with and inhibiting the transcriptional machinery, or indirectly inhibiting transcription by compacting chromatin (Lund and van Lohuizen 2004). Multiple TBP-associated factors (TAFs) have been found to co-purify with fly PRC1, arguing against a pure model of steric hindrance (Saurin et al. 2001). However, it is possible that the localization of PRC1 at promoters is sufficient to prevent transcription by sterically preventing the binding of factors involved in elongation.

One piece of evidence for direct inhibition of transcription comes from an experiment with a *LacZ* reporter driven by an inducible *hsp26* promoter (Dellino et al. 2004). When a PRE is included in cis to the promoter, PcG dependent silencing occurs. Silencing of this construct did not prevent the binding of general transcription factors. TBP and PolII were found to be bound at the promoter, but initiation was blocked. In

addition to possible mechanisms where PRC1 directly inhibits the transcription machinery, it is also possible that the creation of compacted chromatin structure could lead to transcriptional repression.

Compaction of chromatin is another possible mechanism by which PRC1 could potentially silence genes. There are several pieces of evidence that suggest compaction may be a strategy that PRC1 employs. In vitro, proteins from both PRC1 and PRC2 can cause chromatin arrays to condense (Francis et al. 2004; Margueron et al. 2008). The *Drosophila* PRC1 proteins PSC and Su(z)2 compact chromatin in vitro as visualized by electron microscopy (EM), and regions of PSC that are required for compaction are also required in the fly to prevent homeotic phenotypes (Francis et al. 2004; King et al. 2005; Lo et al. 2009). Additionally, we demonstrate here in Chapters two and three that PRC1 proteins from mice, frogs and fish are capable of compacting chromatin. While compaction per se may not prevent the binding of polymerase to promoters, it may block elongation or the activity of nucleosome remodeling enzymes. Remodeling enzymes such as those in the trxG are needed for efficient transcription from certain loci. PRC1 proteins inhibit remodeling in vitro, though it is not known whether this is a result of directly interacting with the nucleosome remodeler or whether it is an indirect result of chromatin compaction (Shao et al. 1999; Francis et al. 2001). Order of addition experiments suggest that for optimal inhibition, PRC1 proteins need to be preincubated with the chromatin substrate prior to addition of remodeler; addition of PRC1 components at the same time as remodeler leads to less efficient inhibition (Francis et al. 2001). This data argue against a model in which PRC1 directly inhibits nucleosome remodeling complexes.

In addition to in vitro experiments, there exists some evidence that PRC1 proteins condense chromatin within the nucleus. In a recent experiment using fluorescent in situ hybridization (FISH) probes, Eskeland et al demonstrated that loss of Ring1B—a PRC1 protein—causes local decompaction of chromatin at *Hoxb* and *d* loci, coincident with an increase of *Hoxb* and *d* gene expression (Eskeland et al. 2010). Reintroduction of wild type Ring1B or catalytically inactive Ring1B leads to rescue. *Hoxb* and *d* gene repression is rescued by either construct, arguing that at these loci, compaction, but not the H2A119ub1 modification, is required for repression. In another study done in *Drosophila*, the genome wide accessibility to the M.SssI DNA methyltransferase was determined (Bell et al. 2010). Lower accessibility, later replication, and lower transcription were observed at H3K27me3 positive sites. Collectively, these studies suggest that chromatin compaction is one mechanism utilized by PcG proteins to silence genes. Precisely how PRC1 compacts chromatin at atomic resolution remains to be seen.

The final question that I will address is how might PcG mediated repression be transmitted through multiple cell divisions. Long-standing unknowns in PcG biology are the mechanisms used to pass the silent state of genes to daughter cells. The propagation of silencing has been proposed to involve the passage of modified histones to the newly replicated DNA. A recent study that analyzed the turnover kinetics of histones using metabolic labeling challenges this proposal. *Drosophila* S2 cells were metabolically labeled with a methionine derivative that allows for ligation to biotin (Deal et al. 2010). This allowed the authors to purify recently deposited nucleosomes and infer kinetics at specific genomic locations based on the DNA sequences that co-purify with newly translated histones. Interestingly, the authors find that nucleosome turnover is quite rapid

(~1.5 hrs) at PSC and ZESTE binding sites, meaning that nucleosomes are turned over multiple times during the cell cycle. Regardless, these turnover rates are slower than those found at sites of active transcription (1 hr). It will be interesting determine whether the turnover kinetics are constant throughout the cell cycle or whether they slow before mitosis. If histone modifications do not transmit epigenetic information to daughter cells, then how might it occur?

Another recent paper analyzes whether the core of PRC1 can remain bound to chromatin templates during replication. In vitro assays demonstrate that PCC can remain bound to nucleosomal templates through replication (Francis et al. 2009). Interestingly, some PCC gets transferred to the newly replicated DNA, suggesting a possible mechanism for epigenetic inheritance. Newly replicated DNA may retain some PcG complexes to maintain the silencing after mitosis.

The different mechanisms discussed above need not be mutually exclusive. Given the large size and diversity of PcG complexes that exist, it seems likely that they use many different mechanisms for repression. Different mechanisms may be utilized at different locations or different tissues.

Dissertation overview

In order to further elucidate PcG-mediated repression we decided to characterize the functions of the core of the mouse PRC1 complex. By analyzing the contributions of different subunits of mouse PRC1 to transcriptional repression we hope to gain a better understanding of the molecular details that contribute to repression. Since PSC is the main engine of repression in *Drosophila* PRC1, the obvious question to address is whether the mouse homolog, Bmi1, is capable of repressive activities. Intriguingly, Bmi1

lacks the extended C-terminal region that is required for in vitro and in vivo activity by PSC.

In chapter two, we explore the mouse PRC1 complex by employing a structure/function analysis to determine whether any of mouse PRC1 proteins is capable of repressive activities. Surprisingly, we find that a protein non-homologous to PSC is responsible for in vitro repressive activities—M33/Cbx2, a homolog of fly PC. Next, we localize the active region of M33/Cbx2 to a part of the protein that is enriched for basic residues and is apparently natively unfolded in solution.

Since it is surprising that a different homolog in the mouse complex contains the repressive activity, and this activity appears to be dependent on charge characteristics, we wondered whether we could make predictions on the functional activity of PcG proteins from other species. In chapter three, a computational approach was used to make predictions about the activity of proteins homologous to PSC—the active fly component—or to M33/Cbx2, the active mouse component. The analysis suggests that sometime during evolution these activities were “swapped” between subunits of PRC1. The predictions were tested by cloning and expressing a panel of proteins from other species and testing whether they can inhibit remodeling and compact chromatin. Finally, in chapter four the implications of the results are discussed and avenues for future research are examined.

Chapter Two: Identification of M33 as a functional homolog of Posterior sex combs

Preface

This chapter is adapted from a Genes and Development paper that was originally published October 15, 2011. It can be examined as published in Appendix A. My contribution to this work was generating Figures 2.1 through 2.9, Figure 2.10B, and table 2.1. Figure 2.10A was generated by Nicole Francis. Joe Garlick expanded the baculovirus used for generating the protein used in far-UV CD spectrometry experiments.

Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge

Daniel J. Grau¹, Brad A. Chapman¹, Joe D. Garlick¹, Mark Borowsky¹, Nicole J. Francis², Robert E. Kingston^{1§}

¹ Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA and Department of Genetics, Harvard Medical School.

² Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138.

Introduction

Protein complexes from the Polycomb group (PcG) group are responsible for the heritable silencing of *Hox* genes during development. Multiple PcG complexes have been characterized. One PcG complex, PRC1, is thought of as the major “engine” of repression. In flies, one subunit from the PRC1 complex, PSC, can inhibit remodeling of nucleosomal arrays as well as compact nucleosomal arrays outside of the context of the complex. However, the precise molecular mechanisms that lead to these activities remain elusive. To further understand how PcG proteins function in general, we undertake a molecular dissection of the mouse core PRC1 complex and compare the functional activities to PSC. By comparing the functional activities of mouse PRC1 to PSC we hope to be able to make biologically relevant generalizations about how PRC1 is able to repress transcription. In this chapter, we demonstrate that a mouse protein, M33/Cbx2—that is not homologous to PSC—is able to inhibit remodeling of nucleosomes. We demonstrate that none of the conserved domains of M33/Cbx2 are required for activity. Using truncated recombinant protein we identify a domain of the protein that is required for optimal activity. Using point mutants that disrupt the positive charge of this domain we demonstrate the importance of charge for optimal inhibition activity. Next we reconstitute core PRC1 complexes with charge mutants of M33/Cbx2 and demonstrate that the charge domain is required in the context of the complex. We provide evidence that this domain is natively unfolded in solution. Finally we provide electron microscopy data that shows that M33/Cbx2 is able to compact nucleosomal arrays.

Results

M33 is a functional homolog of PSC

The *Drosophila* Polycomb group protein PSC is able to block remodeling and to compact nucleosomes in vitro, activities that might directly contribute to PRC1 mediated repression (King et al. 2002; Francis et al. 2004). We hypothesized that if these activities are important to PRC1 function they would be conserved in mammalian PRC1. To investigate this we used the mouse PRC1 core complex (mPCC). As in *Drosophila*, a PRC1 core complex of M33, Ring1A, and Bmi1 retains the majority of repression activity, so we chose to focus on these three subunits, termed here mPCC Δ Ph (Francis et al. 2001; Lavigne et al. 2004). We began by expressing mPCC and individual subunits of the core complex (Figure 2.1.A) to determine which, if any, individual subunits had in vitro activity versus the core.

We characterized the activity of mPCC subunits and subcomplexes using a solution assay that we previously used to characterize PcG proteins. The restriction enzyme accessibility assay (REA) measures the ability of PcG proteins to antagonize nucleosome remodeling by the ATP-dependent remodeling complex hSWI/SNF (Kwon et al. 1994; Francis et al. 2001). Briefly, we assembled nucleosomes onto a 2.5 kB DNA/chromatin template (G5E4) (Figure 2.1.B) using salt dialysis (Utleay et al. 1998). This DNA fragment contains ten 5S nucleosomal positioning sequences. At the center of this fragment is space for two additional nucleosomes, for a total of twelve. With two nucleosomes positioned in the central part of G5E4, one of them occludes a unique HhaI restriction site and prevents cutting. Movement of this nucleosome, which is accomplished efficiently by ATP-dependent nucleosome remodeling, allows HhaI to cut

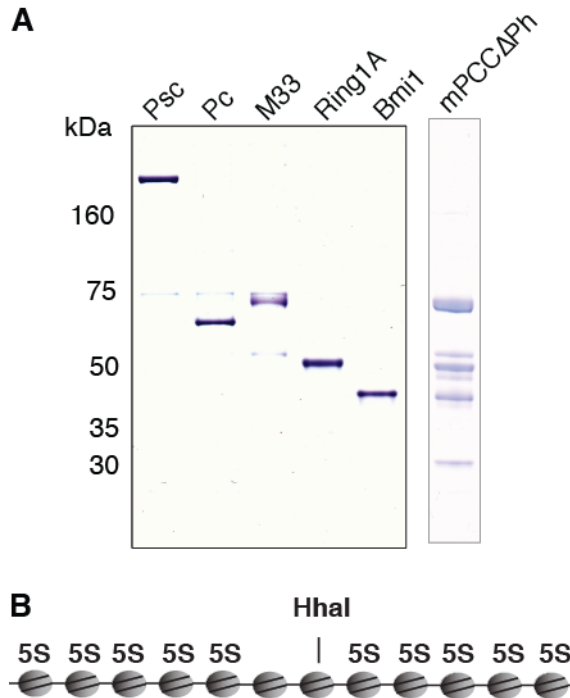


Figure 2.1 Purification of *Drosophila* and mouse PcG proteins

A. Coomassie stained gel of PcG proteins purified from overexpression in *Sf9* cells.

B. Schematic representation of G5E4 nucleosomal array used in assays. 5S: 5S nucleosomal positioning sequence. HhaI: unique HhaI restriction sequence that is inaccessible when packaged around a histone octamer.

the G5E4 DNA. Pre-incubation of the arrays with PcG proteins prevents efficient remodeling of nucleosomes by the hSWI/SNF remodeling complex, and therefore inhibits cutting by HhaI. By titrating in PRC1 proteins or complexes and measuring the amount of cut and uncut DNA we are able to quantify their inhibitory activity.

As anticipated from previous studies, when we preincubate nucleosomal arrays with PSC or with a mouse core complex composed of Bmi1, Ring1A and M33

(mPCCΔPh), we see a concentration dependent inhibition of remodeling (Figure 2.2). Note that this core complex does not contain the Polyhomeotic (Ph) protein. *Drosophila* core complexes with and without Ph have similar in vitro activities (Francis et al. 2001). The inhibition of remodeling activity that we observe with mPCCΔPh was lower than what we had previously measured, which could reflect differences in the activities of protein preparations (Lavigne et al. 2004).

To determine the subunits of mouse PRC1 responsible for inhibition of remodeling activity, we tested proteins individually using this protocol. Surprisingly, we did not observe any appreciable inhibition of remodeling by Bmi1, the mouse homolog of the biochemically active PSC. Likewise, Ring1A did not exhibit any activity. However, M33, the mouse homolog of *Drosophila* PC, exhibited activity that was similar in efficiency to the core complex and to PSC (Figures 2.2.A and B). We conclude that the most active subunit in mouse PCCΔPh for inhibition of remodeling is the M33 protein.

Inhibition seems to require the presence of nucleosomes, as preincubation of M33 with naked DNA template prior to adding HhaI results in inhibition of cleavage that is two orders of magnitude less than when the DNA template is assembled into nucleosomes, suggesting that M33 is not directly interfering with the restriction enzyme (Figures 2.3A and B).

Inhibition of remodeling reactions were performed at ionic concentrations of 60 mM, which are optimal for hSWI/SNF activity (Figures 2.3.C and D). We do not observe significant differences in inhibition of remodeling activity by M33 up to ionic conditions of 175 mM (Figures 2.3E and F).

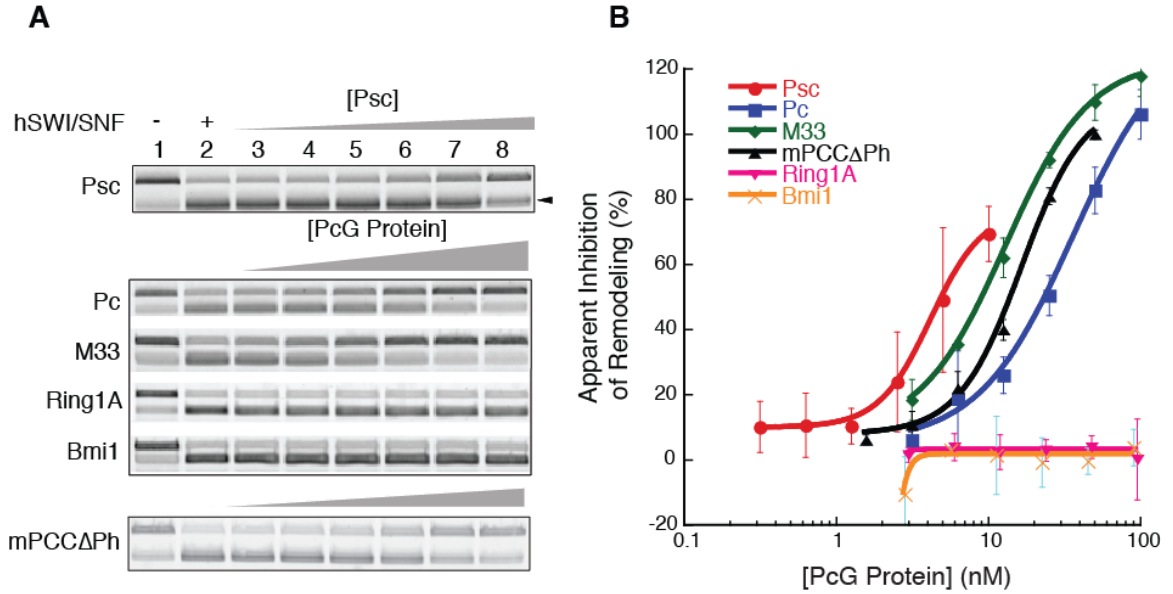


Figure 2.2 M33 is the functional subunit of mouse PCC

A. Agarose gel of restriction enzyme accessibility (REA) assay. After reactions were completed, uncut and cut Cy5 end-labeled G5E4 DNA was separated on an agarose gel, scanned using a Typhoon phosphorimager and quantified using ImageQuant software (GE Healthcare). The slower-migrating band represents the DNA that was not cleaved by HhaI (uncut), while the faster-migrating band represents cut DNA (arrowhead indicates cut DNA).

B. Graph of the data obtained by quantification of DNA bands in panel A. Graphs were created in Kaleidagraph software (Synergy) using a non-linear sigmoidal curve fit. Error bars represent the standard deviation of three technical replicates. Apparent inhibition of remodeling is calculated by the following equation:

$$\frac{(\% \text{Uncut with PcG protein and hSWI/SNF} - \% \text{Uncut with hSWI/SNF})}{(\% \text{Uncut without hSWI/SNF} - \% \text{Uncut with hSWI/SNF})} \times 100$$

These results led us to question the conservation of domain structure between *Drosophila* PRC1 and mammalian PRC1. The *Drosophila* PC protein can inhibit remodeling of nucleosomal templates, but it is at least 5-fold less efficient than PSC, while the mouse homolog of PC (M33) is at least one order of magnitude better at inhibiting remodeling than the mouse homolog of PSC (Bmi1) (Figures 2.2.A and B; note that Bmi1 displays a low level of compaction activity when assessed by electron microscopy, see below). We have previously located the inhibition of remodeling and compaction activities in PSC to its C-terminus, a region with no obvious primary sequence homology to any of the PcG proteins in the mammalian complex. We therefore set out to complete a structure/function analysis of M33 to determine what features of M33 were required for activity, and how those features compared to PSC.

None of the conserved features of M33 are required for in vitro activity

To identify the domain of M33 required for inhibition of remodeling, we expressed and purified M33 and deletion variants in *E. coli*. By using a cleavable GST-tag on the N-terminus and a FLAG-tag on the C-terminus we were able to obtain M33 and variants that were more homogeneous than the proteins we have obtained using the baculovirus system (Figures 2.4A and B). We constructed a series of N- and C-terminal truncation mutants, expressed and purified them, and tested their activity in the REA assay (Figure 2.5A). In agreement with the above results, full length M33 purified from *E. coli* has an inhibitory activity that is similar to M33 purified from *Sf9* cells (compare Figure 2.5B to 2.2B).

concentrations

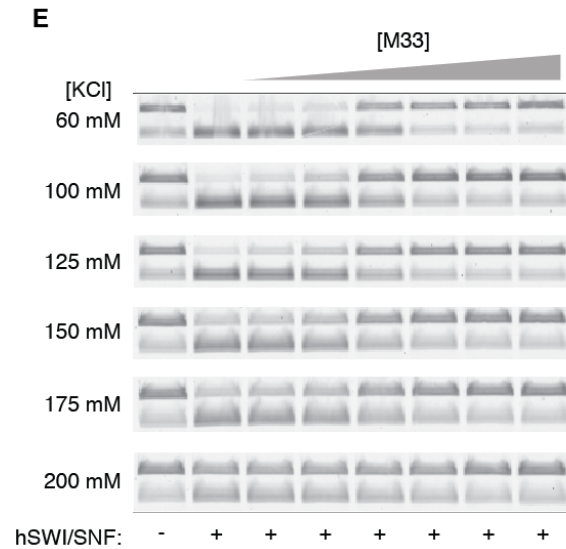


Figure 2.3 (Continued) Inhibition of remodeling requires nucleosomes and optimal salt concentrations

A. REA inhibition activity of M33 preincubated with G5E4 chromatin or naked G5E4 DNA. B. Agarose gel used to generate graph in panel A. For calculating apparent inhibition of remodeling for naked G5E4 it was assumed that 100% of the DNA would be uncut in the absence of HhaI. C. Remodeling activity of hSWI/SNF in the presence of increasing ionic strength. D. Quantification of the results from panel C. E. Inhibition activity of M33 in the presence of increasing ionic strength. F. Quantification of the results in panel E.

M33 does not share significant sequence alignment with PSC, but it contains several domains that are conserved in other chromatin binding proteins: a chromodomain (CHD), an AT-Hook (ATH), and a PC C-box (CBOX) (Figure 2.4A). Another domain, termed here as homology-domain (HD, amino acids 252-266), is a motif that is conserved in M33/Cbx2 homologs but is not found in other Cbx proteins.

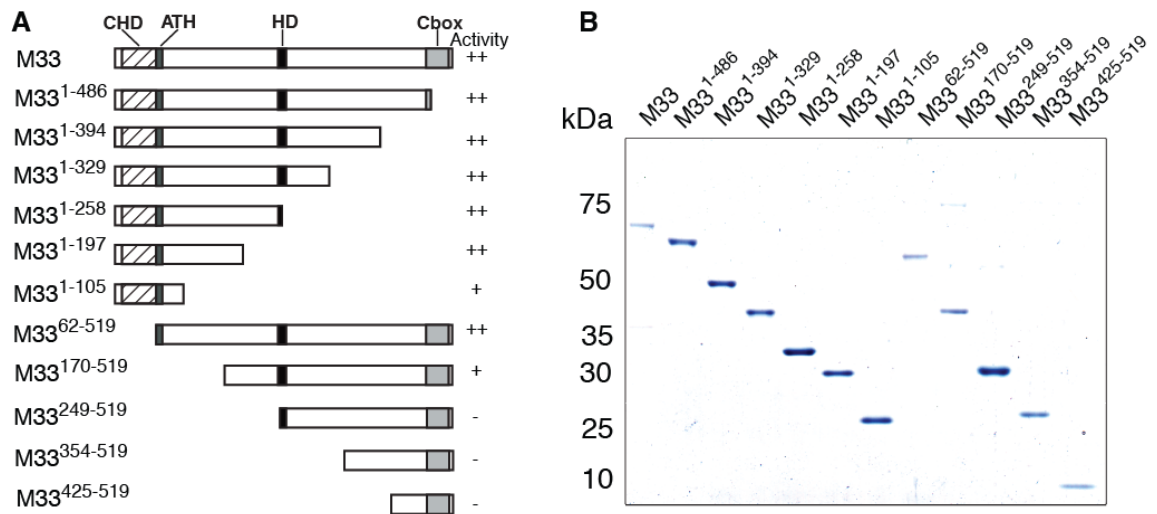


Figure 2.4 Purification of amino and carboxy-terminus truncation mutants of M33

- A. Diagram of M33 truncation mutants tested for biochemical activity. CHD: chromodomain, ATH: AT-Hook motif, HD: homology domain, Cbox: C-box motif.
- B. Coomassie stained gel of the M33 truncation mutants.

We first determined whether any known domains in M33 contributed to its inhibitory activity. Deletion of either the ATH or HD motifs did not result in any reduction in inhibition activity (data not shown). The CHD and CBOX domains interact with nucleosomes, and additionally, the CBOX domain is required for repression activity in *Drosophila* embryos and in transient transfection assays, as well as for interactions with RING proteins (Muller 1995; Schoorlemmer et al. 1997; Breiling et al. 1999; Satijn and Otte 1999; Fischle et al. 2003; Min et al. 2003; Wang et al. 2008). Surprisingly, deletion of the CHD or CBOX domains did not abolish the inhibition activity (Figures 2.5A, B and C).

Since none of the conserved features of M33 are required for our in vitro activity, we wondered whether there might be a non-conserved motif that was responsible for the inhibition of remodeling activity. Thus we decided to create a series of N and C-terminal deletions.

We cloned and expressed a series of truncation mutants of the M33 protein and tested them in the REA assay for inhibition activity (Figures 2.5A, B and C). We find that truncations of the C-terminus to amino acid 105 (M33¹⁻¹⁰⁵) still retain some repression activity. In contrast, N-terminal deletions to amino acid 248 (M33²⁴⁹⁻⁵¹⁹), completely abolishes activity. Since the M33¹⁻¹⁹⁷ and M33⁶²⁻⁵¹⁹ constructs retain near wild type levels of activity—approximately threefold and twofold less, respectively, we conclude that an N-terminal region between amino acids 62 and 197 is required for optimal repression activity.

Figure 2.5 Inhibition of remodeling by M33 truncation mutants

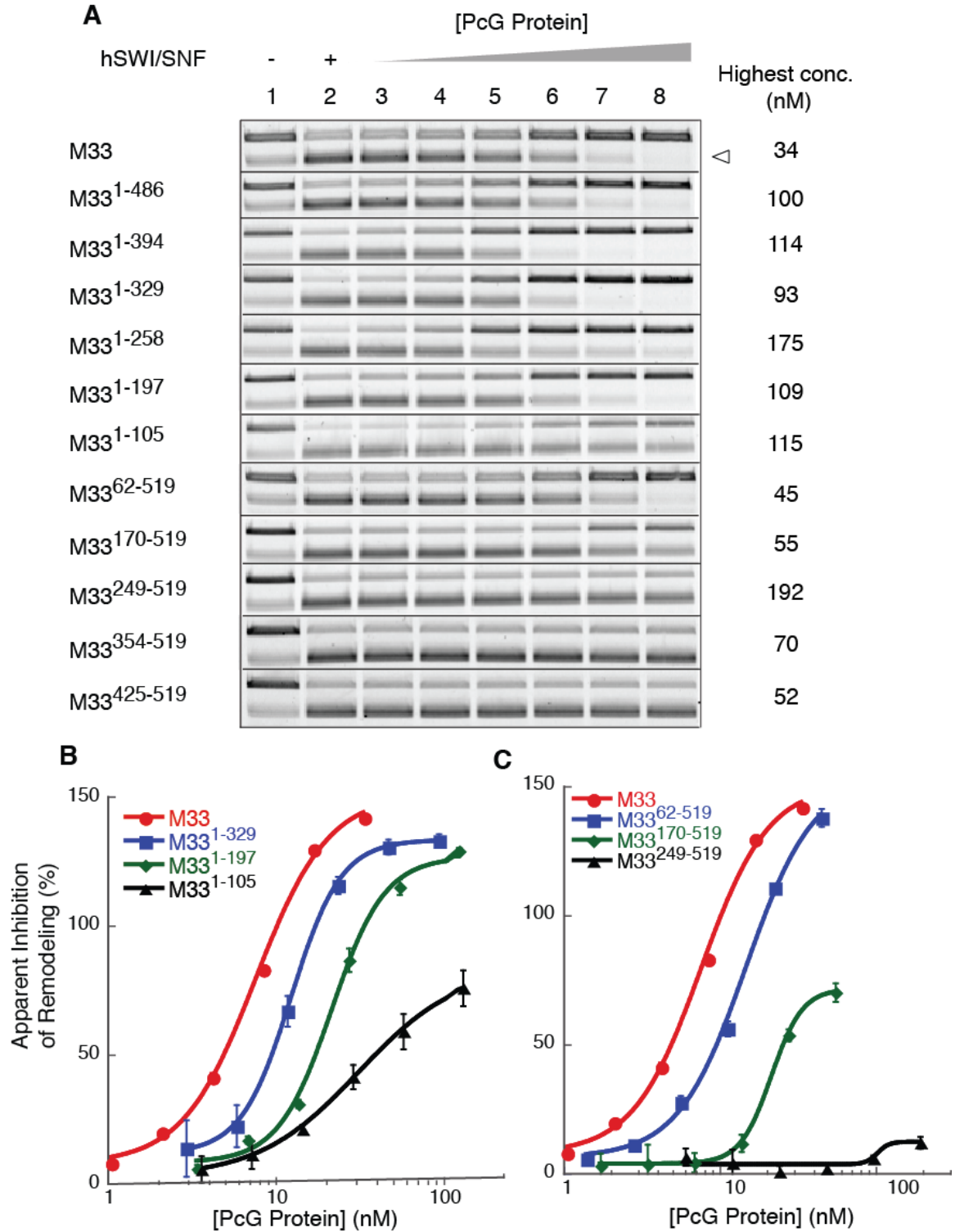


Figure 2.5 (Continued) Inhibition of remodeling by M33 truncation mutants

A. Agarose gel images of the products from the REA assay preincubated with various M33 truncation mutants as described in Figure 2.4A. Numbers represent the highest concentration of variant used in 2-fold dilution series as determined by the Bradford assay and densitometry of the Coomassie stained gel in Figure 2.4B. B. Graph of inhibition activity of selected M33 C-terminal truncation mutants. Data were analyzed as in Figure 2.2. C. Graph of inhibition activity of selected M33 N-terminal truncation mutants.

Positively charged residues are required for optimal activity in vitro

We wondered what characteristics are shared between the regions of M33 and PSC that are required for inhibition of remodeling. Examination of the amino acid composition of M33 revealed a bias towards lysines and arginines--the two most positively charged amino acids. The C-terminal region of PSC that is required for in vitro and in vivo activity also has an overrepresentation of arginines and lysines. The predicted overall charge of M33 and PSC is +32.5 and +82.1, respectively. This is in contrast to Bmi1, inactive in the REA assay, which has a predicted overall charge of +10.1.

We wondered if we could observe any correlation between the predicted overall charge of the M33 variants and inhibition of remodeling activity. Indeed, if we plot the predicted charge of the truncation mutants versus the concentration required for fifty percent inhibition of remodeling, we obtain a linear relationship with $R^2 = 0.7$ (Figure 2.6A). These charged amino acids are spread throughout M33 and PSC, which is reminiscent of another class of repressive proteins: the linker histones (Hansen et al. 2006). We hypothesized that these basic residues in M33 are important for binding to chromatin and therefore contribute to PcG mediated repression. To test this hypothesis we synthesized mutant variants of the M33 protein that perturb overall protein charge. We systematically mutated every other lysine or arginine in M33¹⁻⁴⁸⁶ to alanine, beginning with K132A and ending with K240A, termed here M33^{1-486;1KR_A}, for a total of 13 residues mutated (Figure 2.6B and Table 2.1). We used the M33¹⁻⁴⁸⁶ construct for this

Figure 2.6 Inhibition of remodeling by M33 requires a basic region

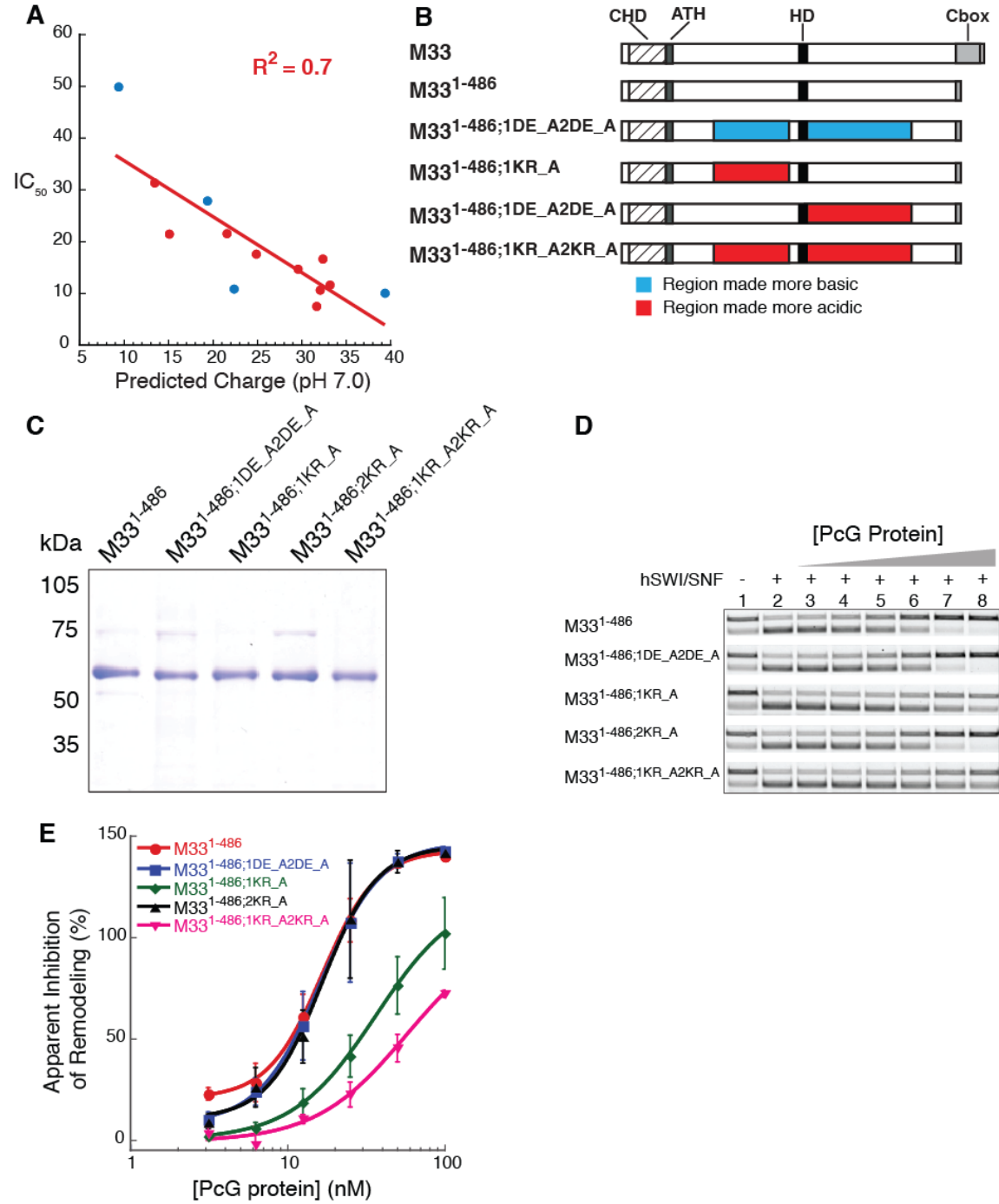


Figure 2.6 (Continued) Inhibition of remodeling by M33 requires a basic region

- A. Graph of predicted protein charge at pH 7.0 versus IC_{50} as determined using Kaleidagraph software and performing a linear fit. Red circles represent data from M33 truncation mutants and blue circles represent data from M33 charge mutants.
- B. Schematic representation of the charge-mutant proteins that were tested.
- C. Coomassie stained gel of M33 charge mutants expressed and purified from *E. coli*. D. Agarose gel with REA reaction products from M33 charge mutants schematically represented in panel B. E. Plot of the quantification from the REA done with M33 charge mutants.

purpose because we find that deletion of the C-terminal 33 amino acids of M33 enhances expression and purification with marginal affect on in vitro activity. Additionally, we generated a construct that mutates every other arginine or lysine from R262A to R415A (M33^{1-486;2KR_A}), as well as combination of the two regions (M33^{1-486;1KR_A2KR_A}). These constructs reduce the predicted overall charge of M33;1-486 at pH 7.0 from +32.5 to +19.3, +22.3, and +9.3, respectively. Finally, we synthesized a construct in which every other aspartic or glutamic acid was mutated to alanine, from E156A to E399A, for a total of 7 mutated residues (M33^{1-486;1DE_A2DE_A}). This construct is expected to increase the predicted charge to +39.3. If overall positive charge of these domains is important to function, these mutations are expected to reduce and enhance the in vitro activity of M33, respectively.

We expressed and purified these proteins to a level similar to M33¹⁻⁴⁸⁶ (Figure 2.6C). When we tested these proteins for inhibition of remodeling activity, we find that M33^{1-486;1DE_A2DE_A} has inhibition activity similar to M33¹⁻⁴⁸⁶ (Figures 2.6D and E). Additionally, M33^{1-486;2KR_A} inhibits remodeling to a similar extent as M33¹⁻⁴⁸⁶. However, both M33^{1-486;1KR_A} and M33^{1-486;1KR_A2KR_A} reduce inhibition activity, approximately 3-fold in the case of M33^{1-486;1KR_A} and 5-fold for M33^{1-486;1KR_A2KR_A} (Data for all of the charge mutants is depicted as blue circles in Figure 2.6A). These results are consistent with the idea that the charged residues within the M33^{1-486;1KR_A} region are important for optimal repression activity and agree with the deletion analysis above that suggests that the region between amino acids 62 and 197 is required for optimal repression activity. However, since there was no enhancement of activity in the M33^{1-486;1DE_A2DE_A} variant,

Table 2.1 Summary of residue changes in M33 charge mutants¹

1-486:1KR A ²	1-486:2KR A ³	1-486:1KR A2KRA ⁴	1-486:1DE A2DE A ⁵
K132A	R262A	K132A R262A	E156A
R136A	K277A	R136A K277A	E173A
K146A	K291A	K146A K291A	E211A
K154A	K296A	K154A K296A	D289A
R162A	K311A	K162A K311A	E319A
K164A	K350A	K164A K350A	D345A
R167A	K366A	R167A K366A	E399A
K175A	K397A	K175A K397A	
R179A	K404A	R179A K404A	
K188A	R415A	K188A R415A	
K192A		K192A	
K216A		K216A	
K240A		K240A	

¹Wild type (1-486) charge: 32.5, pI:10.02

²Charge: 19.3, pI: 9.65

³Charge: 22.3, pI: 9.77

⁴Charge: 9.3, pI: 9.05

⁵Charge: 39.3, pI: 10.27

and no loss of activity in the M33^{1-486;2KR_A} mutant, we conclude that there must be other criteria than simple overall protein charge contributing to inhibition activity of M33.

Since PcG proteins typically function within complexes, we tested whether the results we obtained with M33 alone could be replicated in the context of the core complex. We coexpressed two full length M33 charge variants with Bmi1 and Ring1A in *Sf9* cells and were able to purify both mutant complexes to a level similar to the wild type complex (Figure 2.7A). In agreement with the results obtained with M33 variants alone, PCCΔPh;M33^{1DE_A2DE_A} inhibited remodeling to a similar extent as the wild type core complex, whereas PCCΔPh;M33^{1KR_A2KR_A} reduces activity by more than 4-fold (Figures 2.7B and C).

We wondered whether inhibition activity is a general characteristic of any basic protein. We cloned two mouse genes that contain charge characteristics similar to M33: MrpL2, a mitochondrial protein with predicted charge of +32.96 and CTF8, a nuclear protein with a predicted charge of +30.52. These non-PcG proteins are 7-10 fold less active than M33, suggesting that features beyond overall charge are involved in PcG protein activity (Figures 2.7D, E and F).

To determine whether there are any conserved sequence motifs within the region mutated in the M33^{1-486;1KR_A} construct, we performed a sequence alignment with frog, zebrafish, and chicken Cbx2 proteins (Figure 2.8). We find that five of the mutated residues are conserved among all four species, raising the possibility that these amino acids are involved in nucleosome interactions. While we cannot rule out the loss of activity is due to disruption of protein structure, these results are consistent with the idea that the intrinsic charge of PcG proteins is important for repression activity.

Figure 2.7 Inhibition of remodeling activity of charge mutant PcG

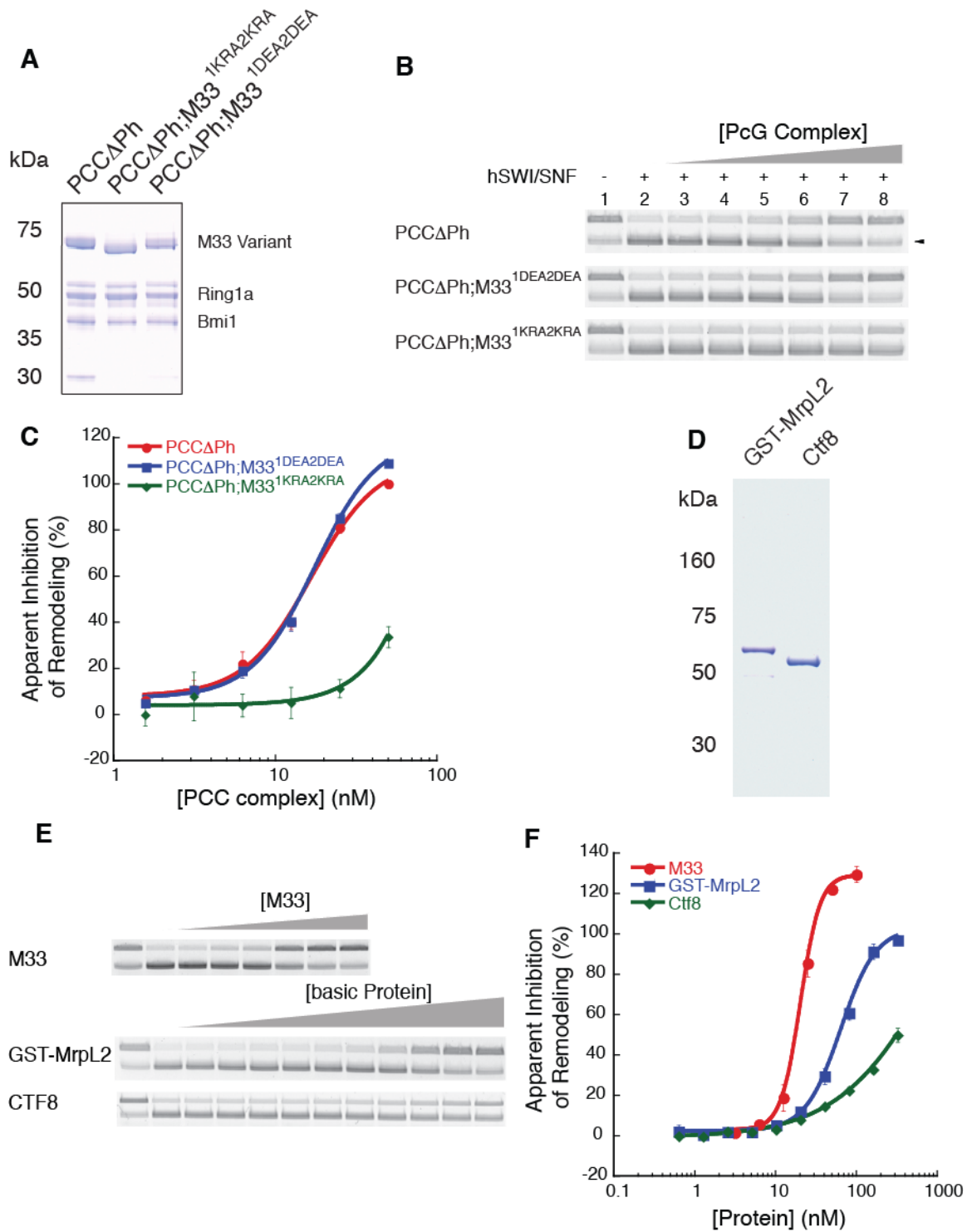


Figure 2.7 (Continued) Inhibition of remodeling activity of charge mutant PcG complexes and basic non-PcG proteins

A. Coomassie stained gel of M33 charge mutants in the context of the core PRC1 complex. Proteins were expressed and purified from *Sf9* cells. B. Agarose gel with REA reaction products from mPCC complexes containing charge mutants of M33. C. Plot of the quantification from the REA done with the charge mutant complexes. D. Coomassie stained gel of the non-PcG basic proteins expressed and purified from *E. coli*. E. Agarose gel with REA reaction products containing non-PcG basic proteins. F. Plot of the quantification from the REA done with the non-PcG basic proteins.

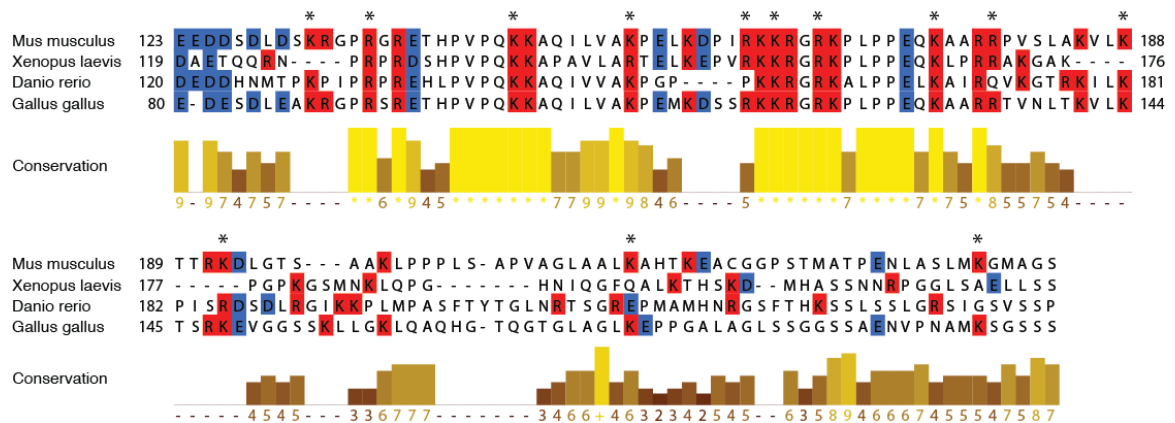


Figure 2.8 Alignment of one of the regions of M33 where basic residues were mutated. Asterisks represent positions where arginines or lysines were mutated to alanine. The alignment was generated by ClustalW and Jalview.

Given the high number of hydrophilic amino acids in M33, one possibility is that the charged region is “natively unfolded” or “intrinsically disordered” (Uversky and Dunker 2010). Natively unfolded regions occur in other chromatin architectural proteins, including the PcG proteins RYBP and GAGA factor, and have been proposed to play a role in the function of PSC and Su(z)2 (Agianian et al. 1999; Emmons et al. 2009; Lo et al. 2009; Neira et al. 2009). The linker histone regions required for chromatin compaction are intrinsically disordered in solution. Genome-wide predictions of natively unfolded regions in proteins forecast a high percentage of transcriptional regulators as having some intrinsic disorder (Garza et al. 2009; Sandhu 2009). Consistent with this idea, Metadisorder, a program that uses several different disorder prediction algorithms, predicts that M33 is folded within the chromodomain and CBOX, while the central portion that contains the 1KR_A region is disordered (Figure 2.9A) (Kurowski and

Bujnicki 2003). Additionally, the Far-UV CD spectra of M33¹⁻⁴⁸⁶ demonstrates that it has the characteristic spectra of a protein that contains intrinsic disorder, specifically, a minima of molar ellipticity at 200 nM (Figure 2.9B) (Kelly and Price 2000).

Compaction of chromatin by M33

These results suggest that in the mammalian PRC1 complex M33 has similar inhibition of remodeling activity to a non-homologous fly protein, PSC, and that this activity corresponds to a basic characteristics of M33. We wondered whether M33 has conserved the ability to compact nucleosomal arrays, an activity that also resides within PSC. To test this, we used electron microscopy to visualize the interaction of mouse PcG proteins with arrays.

PSC can efficiently compact nucleosome arrays, as observed visually by electron microscopy (EM) and as quantified by measuring the end-to-end distance of arrays in the EM images (Francis et al. 2004). We used this technique to examine whether M33 can compact chromatin in a manner like PSC. EM images of PcG proteins preincubated with G5E4 arrays were acquired, and images of low quality were discarded in a double-blind manner. Nucleosomal arrays alone look like the canonical “beads on a string” conformation (Figure 2.10A). Preincubation with Bmi1 tends to reduce the overall array length, but does not appear to promote intranucleosomal interactions. M33, in contrast, promotes the formation of compact particles consisting of multiple nucleosomes. To quantify the ability of these proteins to compact arrays, a single-blind measurement of the end-to-end length of protein/array particles was performed. Full length M33 creates compacted nucleosome structures as determined by a significant decrease in end-to-end

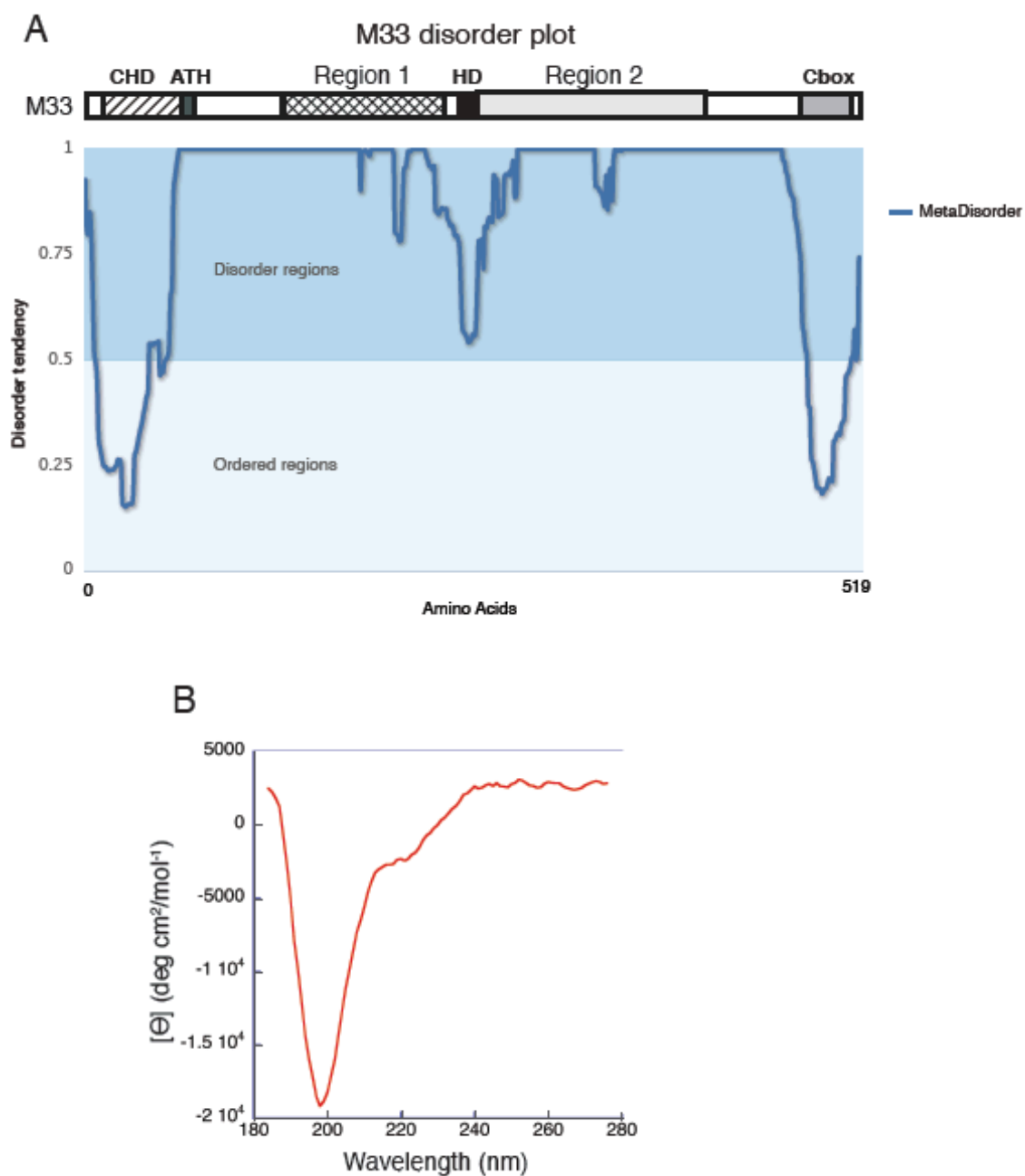


Figure 2.9 M33 and intrinsic disorder

A. Metadisorder plot of predicted protein disorder of M33 with schematic representation of M33 above. Regions over the disorder tendency threshold of 0.5 are classified as disordered. Region 1 and 2 represent the locations where charged amino acids were mutated. B. Far-UV circular dichroism spectra of M33.

length when compared to arrays incubated with no protein (Figure 2.10B, students's T-test, p-value <.0001). Surprisingly, despite being inactive in the REA, Bmi1 is also able to measurably compact arrays. However, M33 is able to compact chromatin significantly better than Bmi1 (p-value <.0001) .

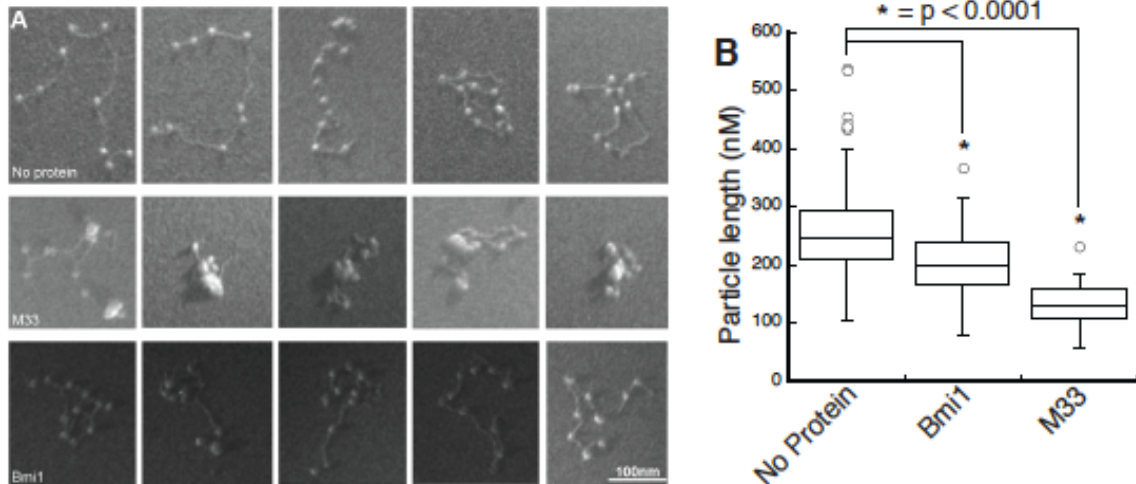


Figure 2.10 Compaction of nucleosomal arrays by mouse PcG proteins

A. Representative EM images of nucleosomal arrays incubated with the indicated PcG protein. B. Box-plot representation of the measured maximal diameter of nucleosomal array particles. Particle length is the diameter of the smallest circle that can entirely surround one nucleosomal array. The box represents the upper and lower quartile, and the line splitting the box represents the mode. The open circles represent outliers, and the asterisks indicate a p-value of < .0001 using Student's t-test. No protein: n=72, Bmi1: n=50, M33: n=30

Materials and Methods

Protein expression

Baculovirus expression was done essentially as described (Francis et al. 2001). Briefly, Sf9 cells were grown in Hyclone CCMIII media at 27°C with shaking. One liter of cells was infected with either individual viruses for the expression of single subunits, or coinfecting with multiple viruses for the expression of protein complexes. After 40 hrs, cells were harvested by centrifugation and washed in PBS prior to making nuclear extracts (NEs) as described (Sif et al. 1998). For the purification of single PcG subunits, NEs were bound to M2 resin (Sigma cat# A2220) for 4 hours, then washed extensively with BC buffer: 20mM HEPES pH 7.9, 0.2mM EDTA, 20% glycerol, 1mM DTT, 10mM PMSF and protease inhibitor cocktail (Roche complete protease inhibitor tablets) containing 500mM KCl. The M2 beads were then washed with BC buffer containing higher concentrations of KCl, up to 2M, before eluting in BC buffer containing 500mM KCl and 0.4mg/ml flag peptide. Purification of PcG complexes was identical except that the bound protein was washed with BC buffer containing 300mM KCl.

For expression of GST fusion proteins in *E. coli*, Rosetta pLysS cells were transformed with pGEX6P1 containing the cDNA of interest. A single colony was used to inoculate 5ml of LB containing 25µg/ml chloramphenicol and 50µg/ml ampicillin, and grown overnight. The following morning, 250µl of the overnight culture was used to inoculate 250ml of the auto-induction media ZYP-5052: 12g/L Bacto tryptone, 24g/L Bacto yeast extract, 25mM (NH₄)₂SO₄, 50mM KH₂PO₄, 50mM Na₂HPO₄, 0.05% glucose, 0.2% alpha-lactose, 0.5% glycerol, 1mM MgSO₄, 25µg/ml chloramphenicol and 50µg/ml ampicillin (Studier 2005). Cultures were shaken at 37°C for 5 hrs, then grown overnight

at 18°C. The cultures were collected by centrifugation at 4,000 RPM in a Beckman J6 MI for 20 min. Cell pellets were resuspended in 40ml of lysis buffer: 50mM HEPES pH 7.5, 0.5mM EDTA, 1.6M KCl, 20% glycerol, 0.5mM MgCl₂, 0.05% NP40, 1mg/ml lysozyme, 1mM DTT and protease inhibitors. The cells were taken through three freeze thaw cycles, then sonicated to shear DNA before centrifugation at 25,000g for 20min to remove debris. 5% polyethylenimine (PEI) in 20mM HEPES pH 7.5 was added dropwise to the supernatant while stirring to a final concentration of 0.15%, and stirred an additional 30min. This step was omitted for GST-MrpL2. The precipitated nucleic acid was removed by centrifugation at 25,000g for 20min. Extracts were bound to glutathione sepharose beads for 2hrs before washing with BC buffer containing increasing salt as described above for M2 purifications. Proteins were either cleaved or eluted off of the resin. For cleavage of proteins, the resin was incubated overnight with 20units of HRV 3C protease in 5ml of BC buffer with 500mM KCl. Proteins were eluted from glutathione sepharose by incubating in BC buffer containing 500mM KCl and 40mM reduced glutathione. Eluted proteins were purified over M2 resin as described above. Purified proteins were quantified using the Bradford assay, and then normalized relative to each other by the intensity of protein bands on an SDS-Page gel analyzed using ImageJ software.

Restriction enzyme accessibility assays

Nucleosome arrays were assembled using HeLa histones as previously described, except that Cy5-labeled G5E4 was used (Sif et al. 2001). Human SWI/SNF was purified from HeLa nuclear extracts as described (Sif et al. 1998). Reactions were carried out in 20µl volume containing 12mM HEPES pH 7.9, 12% glycerol, 60mM KCl, 0.12mM

EDTA, 0.12μg/μl BSA, 2mM ATP, 1.25mM MgCl₂, and 2nM of assembled nucleosomes. Dilutions of PcG proteins were incubated with the arrays for 30min at 30°C prior to the addition of 100ng of SWI/SNF and eight units of HhaI. Reactions were incubated at 30°C for one hour before the addition of 10μl of stop buffer: 1.5mg/ml Proteinase K, 70mM EDTA, 10mM Tris pH 7.7, 1% SDS and 0.1% orange G. Reactions were incubated for 30min at 55°C, then separated on a 1% agarose gel in 1X TAE buffer. DNA was visualized on a Typhoon scanner and quantified using Imagequant software. Apparent Inhibition of remodeling was determined by the following equation: $((\% \text{uncut with PcG and hSWI/SNF} - \% \text{uncut with hSWI/SNF}) / (\% \text{uncut without hSWI/SNF} - \% \text{uncut with hSWI/SNF})) \times 100$, and plotted using kaleidagraph. Curves were generated using an equation for a sigmoidal fit.

Electron microscopy

Electron microscopy experiments were carried out essentially as described previously (Francis et al. 2004), with the following minor modifications. Binding reactions were carried out in either 30 or 60mM KCl, and NP40 added to a final concentration of 0.001%. All reactions were set up at molar ratios of 8 PcG proteins to 1 nucleosome, based upon total protein concentration. This ratio was selected based on electrophoretic mobility shift assays carried out under the same conditions as EM, which demonstrated binding of M33 to nucleosomal arrays with minimal aggregation at this ratio.

Chapter Three: Charge characteristics predicts activity of divergent PcG proteins

Preface

This chapter is adapted from a Genes and Development paper that was originally published October 15, 2011. It can be examined as published in Appendix A. My contribution to this work was generating Figures 3.1 through 3.3 and Figure 3.4B. I generated Table 3.1 based on the bioinformatic analysis done by Brad A. Chapman.

Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge

Daniel J. Grau¹, Brad A. Chapman¹, Joe D. Garlick¹, Mark Borowsky¹, Nicole J. Francis², Robert E. Kingston^{1§}

¹ Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA and Department of Genetics, Harvard Medical School.

² Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138.

Introduction

The results presented in chapter two collectively suggest that a positively charged domain is responsible for the ability of PRC1 proteins to inhibit nucleosome remodeling. In vitro, M33 appears to inhibit remodeling enzymes in a manner that is dose responsive to the overall charge of the protein, whereas PSC appears to be more complicated. Two truncation mutants of PSC that disrupt overall charge to variable degrees do not directly correlate with inhibition activity. PSC¹⁻⁹⁰⁹ contains half as much overall charge as PSC⁴⁵⁶⁻¹⁶⁰³ yet inhibits remodeling to a similar extent (Lo and Francis 2010). This suggests that while overall charge may predict the inhibition of remodeling activity of M33, it may not be a characteristic that can be used to make generalizations for all proteins. Indeed, two proteins that have high overall charge—MrpL2 and Ctf8—were not as active as M33 in inhibition assays.

Since M33 and PSC are not sequence homologs, we hypothesized that these proteins evolved to become functional homologs. This hypothesis predicts that organisms that are more closely related to *Drosophila* will have PSC homologs that are active. Likewise, it predicts that organisms more related to the mouse will have M33 homologs that are classified as active. In chapter three, we test these predictions by using a computational approach to classify proteins and then test these predictions using functional assays.

Results

Evolution of PcG proteins suggests charge is a predictor for in vitro activity

In order to make predictions about the in vitro activity of PRC1 proteins we first generated a list of homologous proteins from other organisms. Homologous M33/PC and

Bmi1/PSC proteins were identified by searching the UniProt database for conserved chromodomain (IPR000953) and RING (IPR001841) InterPro signatures (Hunter et al. 2009; 2011). Identified proteins from this search were filtered to PcG-like proteins using additional regions of homology: the CBOX domain for M33/PC and the extended RING homology domain for Bmi1/PSC by using hmmsearch (Eddy 2009).

Using PSC and M33 as known active proteins, and Bmi1 as a known inactive protein, the PcG proteins were classified as active or non-active using k-means clustering (Cock et al. 2009). The clustering was based on overall protein charge and regional charge. Overall charge was calculated using Biopython and regional charge was calculated as the percentage of 75 amino acid windows with an isoelectric point (pI) greater than 10.2. Proteins of the active class contained a regional charge of at least ten percent and overall charge of at least +15. Additional classification parameters, such as protein interaction and domain distribution, were considered but excluded as non-informative.

The total number of chromodomain and RING domain proteins identified were 44 and 59, respectively, from 13 species (Table 3.1). Of these proteins, 32 chromodomain proteins and 4 RING proteins are predicted to inhibit remodeling. We then used 18S ribosomal RNA (rRNA) sequences from these 13 species to generate a phylogenetic tree to show the evolutionary relationships of these organisms (Figure 3.1). Interestingly, when we did this we observe an evolutionary point where the predicted activity of RING-containing PcG proteins and chromodomain-containing proteins diverged. In deuterostomes, the PcG proteins predicted to be active are all chromodomain-containing. In protostomes, only predicted active RING domain containing proteins are observed.

Notably, despite the expansion of RING proteins in deuterostomes, we could not observe any Bmi1/PSC homologs that the analysis predicts to be active. We used this list to test the accuracy of our predictions.

To accomplish this, we expressed and purified six evolutionarily divergent chromodomain and RING containing PcG proteins and tested their activity in vitro.

These proteins were from the species *Xenopus laevis*, *Danio rerio*, and *Caenorhabditis elegans* (Figures 3.2A and B). Overall, we tested one predicted inactive RING protein: frog Pcgf2, one predicted inactive Cbx protein: zebrafish Cbx7, one predicted active RING protein: worm Mig-32, and three predicted active Cbx proteins: zebrafish Cbx8, frog Pc1, and Cbx6 (see Figure 3.2A for predicted charge).

Proteins were chosen based on evolutionary divergence from either mouse or *Drosophila*, and the availability of cDNAs. When we tested the activity of these proteins using the REA assay we find that, as predicted, frog Pcgf2 and zebrafish Cbx7 do not inhibit remodeling activity (Figures 3.2C and D). Conversely, preincubation with arrays using worm Mig-32, frog Pc1 or Cbx6, or zebrafish Cbx8 leads to efficient inhibition of remodeling. We do not observe any effect from leaving the GST-tag on our proteins nor for GST alone (Figure 3.3). Thus, for these proteins, regional charge was an accurate predictor of in vitro activity.

Table 3.1 Classification of predicted active and inactive PcG proteins**Predicted active Cbx proteins**

Org.	Uniprotkb ref#	Other names	Charge ¹	Regional Charge ²
M. musculus	O55187	Cbx4,Pc2	25.78	0.17
	P30658	Cbx2,M33	32.51	0.36
	Q6TKP2	Cbx6,Npcd,Nptxr	18.58	0.28
	Q9DBY5	Cbx6	23.57	0.5
	Q9QXV1	Cbx8,Pc3	19.7	0.32
R. norvegicus	Q4FZS1	Cbx8	18.71	0.3
	Q5M9G4	Cbx6	22.33	0.48
H. sapiens	B0QXZ6	CBX6,RP3-327J16.5-003	23.85	0.49
	O00257	CBX4	27.08	0.17
	O95503	CBX6	25.09	0.51
	Q14781	CBX2	35.24	0.41
	Q9HC52	CBX8,PC3,RC1	19.69	0.29
B. taurus	A3KMW6	CBX6	22.33	0.5
	Q0P596	CBX8	18.68	0.28
B. floridae	B6LD55	BRAFLDRAFT_118973	25.29	0.47
	B6LD70	BRAFLDRAFT_118976	24.72	0.39
D. rerio	A0JMJ7	cbx8	20.6	0.22
	A8KC45	zgc:111978	21.28	0.28
	Q1LWE1	DKEY-205O12.3-001,zgc:101049	37.49	0.49
	Q5U3G5	cbx2	32.33	0.44
	Q804D1	cbx4	20.68	0.11
X. laevis	Q5M982	LOC496172	24.36	0.46
	Q63ZN4	LOC494770	16.07	0.44
	Q6NTQ8	MGC83092	25.61	0.3
	Q91647	cbx4, Xpolycomb	15.8	0.1
	Q9YHC4	Pc1	27.29	0.54
N. vectensis	A7RJQ9	v1g238789	16.3	0.19
G. gallus	O93482		16.49	0.25
T. nigroviridis	Q4RGD8	GSTENG00034849001	32.26	0.58
	Q4RJJ7	GSTEBG00033406001	30.65	0.1
	Q4SAU8	GSTEBG00021254001	18.66	0.64
	Q4T358	GSTEBG00007993001	28.97	0.58

Table 3.1 (Continued)**Predicted inactive Cbx proteins**

Org.	Uniprotkb ref#	Other names	Charge	Regional Charge
M. musculus	Q8VDS3	Cbx7,D15Ert417e	-4.76	0
R. norvegicus	P60889	Cbx7	-3.76	0
H. sapiens	B0QYP2	CBX7,RP4-742C19.7-003	-3.76	0
	O95931	CBX7	10.15	0.3
	Q6TPI8	O00257-2	4.78	0.25
B. floridae	B6LD56	B6LDRAFT_118974	10.84	0.28
D. rerio	Q568R0	zgc:110152	-7.73	0.22
N. vectensis	A7S7R4	v1g243415	-2.89	0
T. nigroviridis	Q4RJJ6	GSTENG00033407001	10.22	0.09
	Q4S4B5	GSTENG00024249001	1.94	0.21
	Q4SVN4	GSTENG00011899001	12.08	0.42
D. melanogaster	P26017	Pc,CG7618	4.22	0

Predicted active RING proteins

Org.	Uniprotkb ref#	Other names	Charge	Regional Charge
C. elegans	Q19336	mig-32,F11A10.3	36.66	0.31
A. gambiae	A0NES3	AGAP006403	86.68	0.32
D. melanogaster	P25172	Su(z)2,CG3905	46.9	0.3
	P35820	Psc,CG3886	82.12	0.31

Predicted inactive RING proteins

Org.	Uniprotkb ref#	Other names	Charge	Regional Charge
M. musculus	ASASR8	Bmi1,RP23-396N6.2-003	2.2	0
	B2RTD3	Pcgf5	0.11	0
	P23798	Pcgf2,Mel-18,Mel18,Rnf110,Zfp144,Znf144	4.56	0.01
	P25916	Bmi1,Bmi-1,Pcgf4	10.1	0
	Q3UK78	Pcgf5,Rnf159	-1.02	0
	Q6PHA8	Al451617	-11.19	0.09
	Q8BTQ0	PcgF3,Rnf3,Rnf3a	7.28	0.04
	Q8R023	Pcgf1,Nsps1,Rnf68	12.82	0.07

Table 3.1 (Continued)**Predicted inactive RING proteins**

Org.	Uniprotkb ref#	Other names	Charge	Regional Charge
M. musculus	Q99NA9	Pcgf6,Mblr,Rnf134	-16.11	0
R. norvegicus	B2RZ26	Pcgf3	7.28	0.04
	B2RZ82	Pcgf2,Pcgf2_predicted,rCG_32972	4.56	0.01
	B2RZ90	LOC687730	0.35	0
	B4F7B6	Bmi1,Pcgf4_predicted,rCG_55775	11.1	0
	Q5XI70	Pcgf6,Rnf134	-17.11	0
	Q6DLV6	Pcgf1,Nspc1	12.58	0.07
H. sapiens	B4DTN5		14.88	0.16
	P35226	BMI1,PCGF4,RNF51	10.1	0
	P35227	PCGF2,MEL18,RNF110,ZNF144	4.59	0.03
	Q3KNV8	PCGF3,RNF3,RNF3A	7.52	0.04
	Q5T8Z1	BMI1,RP11-573G6.1-004	2.2	0
	Q5T8Z2	BMI1,RP11-573G6.1-005	8.7	0
	Q86SE9	PCGF5,RNF159	-1.98	0
	Q9BSM1	PCGF1,NSPC1,RNF68	12.82	0.07
	Q9BYE7	PCGF6,MBLR,RNF134	-20.11	0
B. taurus	A0JN86	PCGF5	0.02	0
	Q2KJ29	PCGF3	7.52	0.04
	Q2YDF9	PCGF1	12.3	0.01
	Q32KX7	PCGF4	10.1	0
B. floridae	B6LC96	BRAFLDRAFT_118878	-1.59	0.01
	B6MQT0	BRAFLDRAFT_85508	7.76	0
	B6NBB8	BRAFLDRAFT_93084	6.58	0
	B6NBC6	BRAFLDRAFT_126892	-1.22	0.08
	B6P4H7	BRAFLDRAFT_242708	-2.35	0
D. rerio	A2ATX3	si:ch211-67n3.7,CH211-67N3.7-001	-0.65	0
	B3DK16	LOC562327	4.8	0
	Q1JPS1	pcgf5,zgc:136815	0.56	0
	Q7T3E6	bmi1b,pcgf4b	8.07	0.04
	Q7ZYZ7	pcgf1,si:zc207o21.2	12.33	0.09
	Q8JIR0	bmi1a,bmi1,pcgf4a,psc1	8.07	0
X. laevis	Q4QR06	pcgf1	15.09	0.08
	Q6NTM9	MGC83173	8.62	0.01
	Q6NU75	pcgf2,MGC81178	5.38	0
	Q91648	bmi1a,pcgf4,xbmi-1	10.58	0

Table 3.1 (Continued)**Predicted inactive RING proteins**

Org.	Uniprotkb ref#	Other names	Charge	Regional Charge
N. vectensis	A7SGN9	v1g53309	6.07	0
	A7SGP3	v1g117700	10.54	0.04
	A7SGP4	v1g117725	8.21	0
G. gallus	Q5SDR3	BMI1,PCGF4	10.1	0
T. nigroviridis	Q4RZN6	GSTENG00026396001	8.59	0.05
	Q4SG67	GSTENG00018789001	2.1	0
	Q4SN13	GSTENG00015488001	4.89	0.01
	Q4SWB5	GSTENG00011573001	1.49	0.01
A. gambiae	Q7QCA8	AGAP002533	5.03	0.03
D. melanogaster	Q4V707	Su(z)2	2.12	0.08
	Q9TXC7	Su(z)2,CG3905	-7.25	0
	Q9VV77	l(3)73Ah,CG4195,Dmel_CG4195	8.04	0.03

¹ Charge is overall predicted charge determined by using Biopython.

² Regional charge is the fraction of sliding windows of 75 amino acids with an average isoelectric point of 10.2 or higher.

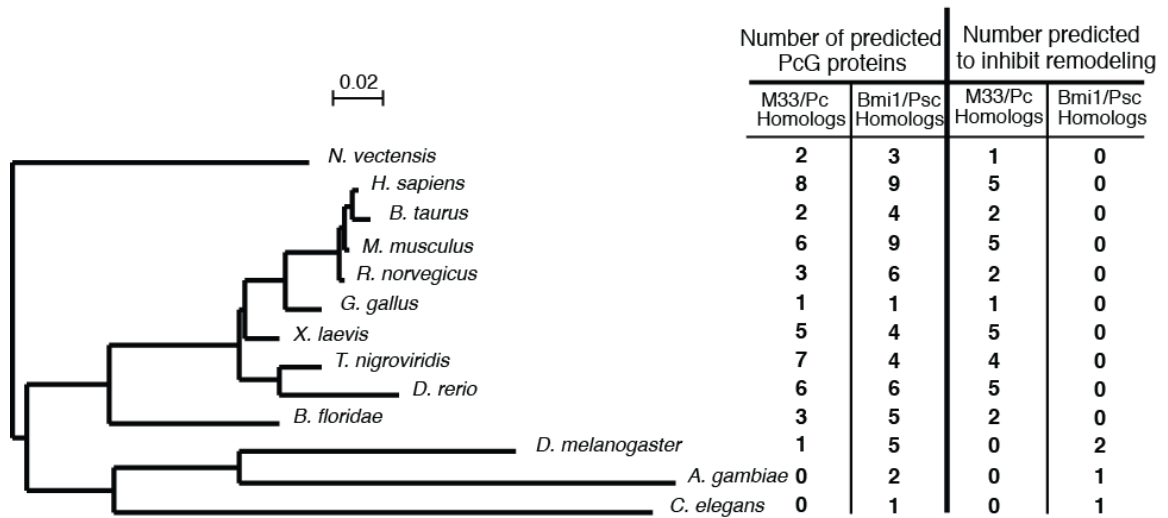


Figure 3.1 Analysis of evolutionary conservation of PcG function

Phylogenetic tree of species containing RING domain or Chromodomain proteins from UniProtKb protein database. The tree is based on alignments of 18S rRNA from each of the species. Number of predicted PcG proteins represents the number of each class of proteins that was found in the UniProtKb database. Number predicted to inhibit remodeling is the number of proteins from each class that is expected to have inhibition activity based on overall protein charge and regional charge. The scale bar represents 0.02 substitutions per site.

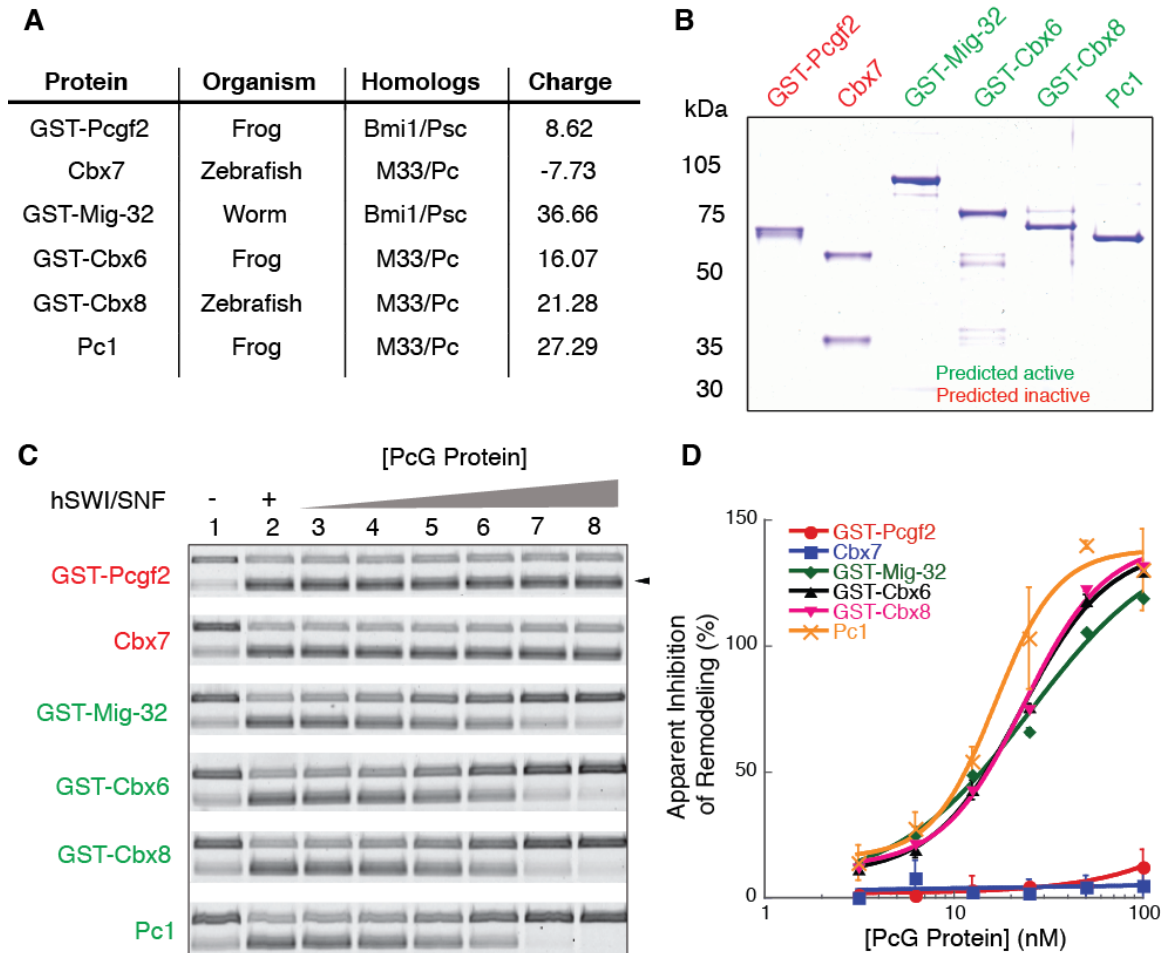


Figure 3.2 Inhibition of remodeling activity of PcG proteins from other species

- A. Charge properties of PcG proteins selected for in vitro activity analysis. The accession numbers for the proteins are: Pcgf2: NP_001084738.1; Cbx7: NP_001017853.1; Mig-32: NP_502293.2; Cbx6: NP_001088074.1; Cbx8: AAI54356.1; Pc1: NP_001081900.1.
- B. Coomassie stained gel of PcG proteins expressed and purified from *E. coli*.
- C. Agarose gel of REA assay reaction products.
- D. Plot of the quantification of results obtained in panel C.

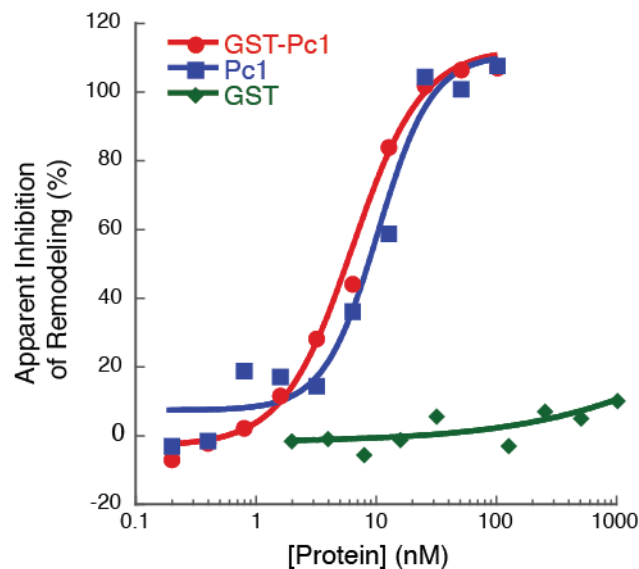


Figure 3.3 GST does not impact inhibition of remodeling activity

Inhibition of remodeling activity of GST-tagged or untagged Pc1 and GST alone. A ten-fold molar excess of GST did not inhibit remodeling.

Compaction of chromatin by divergent PcG proteins

Our results above are consistent with the hypothesis that the domain primarily responsible for inhibition of remodeling *in vitro* resides on the PSC homolog in flies and worms and resides on the PC homolog in vertebrates. This idea posits that one subunit of PRC1--either an M33/PC homolog or a Bmi1/PSC homolog--will contain a region (or regions) of high positive charge that is important to repression of remodeling and also for compaction. As we had previously done with M33 and Bmi1, we tested whether the proteins from species other than fly and mouse are able to compact nucleosomal arrays using electron microscopy

We quantified the ability of the PcG proteins from divergent species to compact arrays. In agreement with the previous REA assay, neither GST-Pcgf2 nor Cbx7 (predicted inactive) is able to significantly compact arrays (Figures 3.4A and B, p-values 0.1 and 0.8 respectively). These reactions yield arrays that have extended conformations. However, the proteins that inhibit remodeling (GST-Mig32, GST-Cbx6, GST-Cbx8 and Pc1) are able to significantly compact arrays (p-value < .0001) (Figures 3.4A and B). As with M33, these proteins promote intranucleosomal interactions.

Thus, we have predicted PcG protein activity based on regional charge and overall charge and have shown that the predictions appear to hold true. Proteins that we predict to be active both inhibit remodeling and compact nucleosomes in vitro. In contrast, the predicted inactive proteins do not.

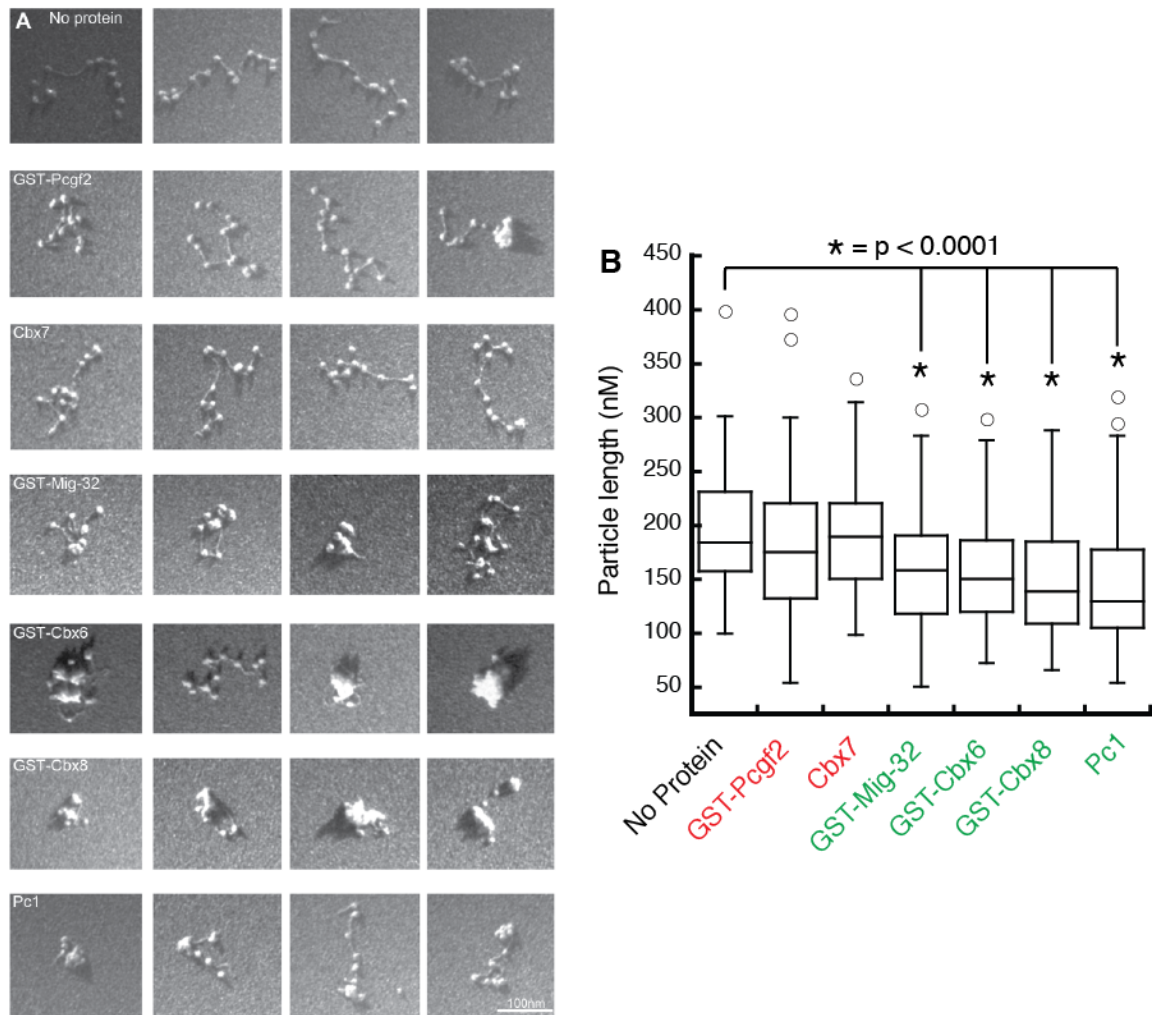


Figure 3.4 Compaction of nucleosomal arrays by diverse PcG proteins

A. Representative images of nucleosomal particles incubated with various PcG proteins from different species.

B. Box-plots of images as described above. No protein: $n=79$, GST-Pcgf2: $n=86$, Cbx7: $n=113$, GST-Mig-32: $n=90$, GST-Cbx6: $n=88$, GST-Cbx8: $n=87$, Pc1: $n=89$.

Materials and Methods

Computational classification of PcG proteins

The UniprotKb database was queried for proteins containing chromodomains (IPR000953) or RING domains (IPR001841). This list was filtered for PcG proteins using HMMER by keeping only those proteins containing a CBOX for chromodomain proteins or extended homology for RING proteins. The predicted charge and regional charge of known active and inactive PcG proteins (regional charge is defined as the percentage of sliding windows of 75 amino acids where the average pI is 10.2 or above) was used to cluster the unknown PcG proteins into either active or inactive classes using the k-means method. Proteins of the active class contained at least 10% regional charge and overall charge of 15 or greater. The reproducible Python scripts for this analysis are available: https://github.com/chapmanb/mgh_projects/tree/master/dg_PSC.

The phylogenetic tree was generated by using ClustalW to create an alignment of 18S rRNAs, and then using the maximum-likelihood method. The tree was drawn using NJplot.

Protein expression

Same as in chapter two.

Restriction enzyme assay

Same as in chapter two.

Electron microscopy

Same as in chapter two.

Chapter Four: Discussion and conclusions

Overview

In chapter two we have shown that the predicted protein charge of a mouse PcG protein correlates with in vitro activity. We extended this observation in chapter three by making a computational prediction of PcG activity in a variety of species and demonstrated that we can predict activity based on charge characteristics. These results support the hypothesis that one key function for PRC1 proteins is the ability to compact nucleosomal arrays and repress chromatin remodeling. The conservation of this basic, charged domain suggests it may be important to silencing by PRC1 family proteins.

Roles of natively unfolded proteins

Natively unfolded or intrinsically disordered proteins were first described in the late 1980s (Sigler 1988). These early descriptions were focused on the proteins that are involved in transcriptional activation. Notably, it was observed that the negatively charged amino acids of proteins required for optimal transcriptional activation did not need to be precisely ordered; Ma and Ptashne elegantly demonstrated that the critical parameter appeared to be amino acid composition (Ma and Ptashne 1987). We find that canonical transcription repressors, the PcG proteins, also appear to have regions of disorder, yet, in contrast to transcriptional activators, contain high concentrations of basic amino acids. It is tempting to speculate that these oppositely charged disordered regions play a “yin-yang” role in transcriptional regulation. It is possible that, in addition to the roles in nucleosome interaction described above, these positively charged transcription repressors could directly interact with and inhibit the negatively charged activation domains of the transcriptional machinery.

There are several proposed reasons why proteins would contain regions of disorder. Disordered regions could potentially adopt different conformations that allow interactions with multiple binding partners. This “hub” function is expected to be beneficial for regulatory proteins; a single protein could potentially regulate many different proteins in a context specific manner (Gunasekaran et al. 2003). There is also the “fly-casting” model, where an extended conformation could allow a protein to “sample” a larger amount of space, forming and breaking low affinity contacts until conformational change induces tighter binding (Shoemaker et al. 2000). This is expected to promote interactions of low affinity and high specificity. One computational predictor of protein disorder—charge—was found to also be predictive of PcG functional activity, suggesting that charged disordered regions could possibly play a general role in PcG mediated repression.

Charged domains and PcG function

What might be the biological role for PcG charged domains in the repression of transcription? They appear to be predictive for both inhibition of remodeling and compaction of chromatin in vitro. Here we propose a model for how the charged domains of PRC1 function: 1) PRC1 is recruited to target loci and presents the charged domain to linker and/or nucleosomal DNA (Figure 4.1A). 2) The charged domain initially interacts with a nucleosome and creates more interactions with other nucleosomes (Figure 4.1B). 3) Finally, oligomerization occurs through Ph or other protein-protein interactions to promote spreading or formation of higher order chromatin fibers. (Figure 4.1C).

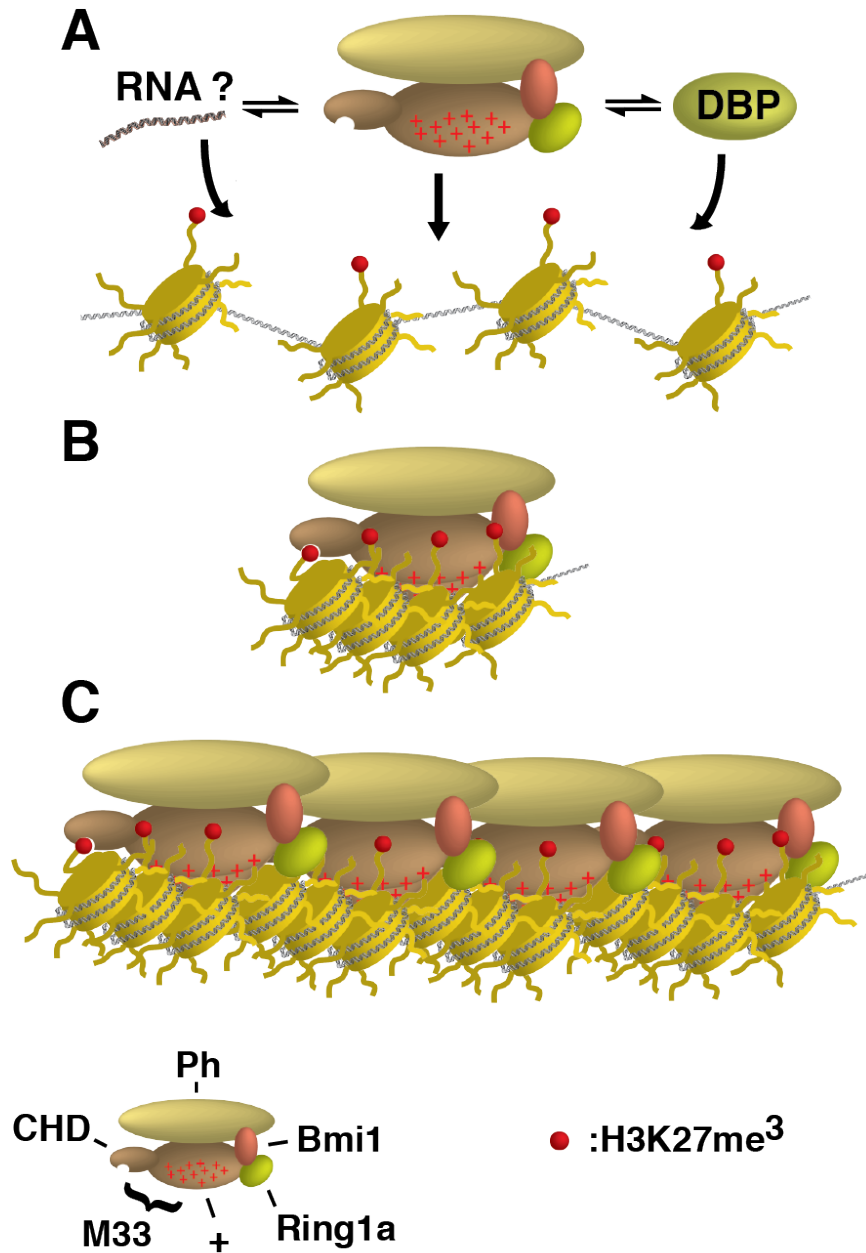


Figure 4.1 A model for chromatin compaction by the mouse PRC1 core complex

A. The mouse PRC1 core complex is recruited to target loci, potentially through a variety of mechanisms. DBP: DNA binding protein. B. The charged region of M33 (indicated by plus signs) interacts with nucleosomes to compact chromatin. C. Further protein-protein interactions from other proteins in the core PRC1 complex drive spreading of compacted chromatin.

The CBOX domain of M33 is not required for in vitro repression activities, yet this motif is conserved and required for the repression of template DNA in cell-based assays (Schoorlemmer et al. 1997). This domain is required for interactions with Ring1A/B and Bmi1, which in turn interact with Ph proteins (Alkema et al. 1997; Gunster et al. 1997; Hashimoto et al. 1998; Hemenway et al. 1998; Satijn and Otte 1999). Thus we imagine an initially transient nucleosome-nucleosome interaction mediated by charged domains facilitates the further stabilization of a repressed chromatin structure that is mediated by other PRC1 proteins. Studies analyzing the dynamics of Cbx/chromatin interactions in culture cell models observe both transiently and stably associated Cbx proteins, consistent with an initial unstable interaction followed by step(s) that promote stable associations (Ren et al. 2008).

PcG protein evolution

What might be the explanation for how the charged domain evolved to reside on an M33/PC homolog in mammals as opposed to a Bmi1/PSC homolog in flies? We can imagine at least four possible explanations: 1) An early common ancestor had both a M33/PC and Bmi1/PSC homolog that each contained a charged region; sometime during evolution the characteristic charge was lost from one or the other homolog. 2) The charged region initially resided on either an M33/PC homolog or a Bmi1/PSC homolog in the common ancestor, but was lost during evolution and subsequently gained on the other PcG homolog. 3) The charged region initially resided on either an M33/PC homolog or a Bmi1/PSC homolog in the common ancestor, was gained on the other before being lost. 4) A convergent evolutionary event occurred: neither the M33/PC nor

the Bmi1/PSC homolog in the common ancestor had a charged domain, but during evolution one homolog or the other evolved it.

Resolution of the possibilities mentioned above will require a more extensive examination of potential PcG members across evolution than has been performed here. Our phylogenetic tree was not a comprehensive list of all deuterostomes and protostomes, so it is possible that there may be unexpected active homologs in other species. Additionally, protein sequences may have been incorrectly annotated in the database that we queried or missing from organisms with incomplete genomes. One unclear example involves Su(z)2, a homolog of PSC that can inhibit remodeling in vitro. Su(z)2 was found in both the predicted active and inactive lists, although the isoforms in the inactive list appear to be short fragments of the full length protein that may not exist in vivo. Regardless of a full examination of this issue, the data reported here are consistent with the hypothesis that there was evolutionary pressure to maintain a highly charged domain in a component of PRC1. One possibility is that across multiple species the core of PRC1 will contain several conserved domains/motifs: two RING fingers, a chromodomain, a SAM domain and a disordered/charged domain--defining the basic functional unit of PRC1.

Molecular nature of PcG/chromatin interactions

We do not understand the precise molecular mechanisms behind PcG protein interactions with chromatin. The flexible charged domains might interact with linker DNA, with nucleosomal DNA, with the histones themselves, or a combination of these chromatin components. We found that two non-PcG proteins with predicted charges similar to M33 do not inhibit remodeling as well as M33. This suggests a mechanism that

does not rely solely upon the amount of positive charge. It is possible that function involves a specific spacing of the charged residues and/or juxtaposition of the charged surface with other functional domains. For example, the majority of the active proteins that we characterized also contain a chromodomain, a known histone-binding domain, opening up the possibility that both DNA and histone contacts are required for optimal PcG repressive activities.

Previous studies have shown that intrinsically disordered regions of proteins can become folded upon interacting with their substrate. This gives us hope that eventually the molecular mechanisms of chromatin condensation by PcG proteins can be unraveled using structural approaches.

Technical challenges overcome

Perhaps the greatest challenge to overcome regarding the biochemical study of PcG proteins is preparing the proteins. This technical challenge can be further subdivided into two separate issues: the expression of proteins at high enough levels to allow purification, and the subsequent purification of those proteins. The baculovirus expression system has been used to express many proteins that otherwise cannot be expressed in bacterial systems. I initially began expressions for the M33/Cbx2 variants and the evolutionary proteins using the baculovirus system. Full length M33 often was poorly expressed, and affinity purifications led to low amounts of purified protein, (~100 ug/L expressed) of relatively low quality (~30% pure via Coomassie stain). Subsequent attempts to further purify M33/Cbx2 using ion exchange chromatography and size exclusion chromatography resulted in further loss of protein.

Similar results were obtained when expressions of the PRC1 proteins from frog, fish, and worm were attempted.

To attempt to overcome the limitations that were being faced using the baculovirus system, I chose to optimize the expression of soluble PcG protein using *E.coli* as an expression host. In ideal cases, expression in bacterial systems can lead to many milligrams of protein being produced per liter of cells cultured. Initial attempts at expression gave low yields of soluble protein. The rapid induction of protein synthesis by IPTG induction can often lead to the formation of inclusion bodies—insoluble aggregates of recombinant protein. Two strategies that have been used to increase the expression of soluble protein in *E. coli* are growth at lower temperature and slower induction. To this end, cultures were grown in auto-inducing media (AIM) overnight at 18 °C.

The phenomena of auto-induction in bacteria was observed by F.W. Studier when he observed the induction of protein expression in cultures approaching saturation in the absence of inducer (Studier 2005). He hypothesized that low levels of lactose in the media were responsible for inducing the *T7lac* promoter when the cultures were nearing saturation, and reasoned that it was a result of the mechanisms bacteria use to regulate the use of carbon sources. At low densities, bacteria preferentially use glucose and uptake of lactose is prevented (Meadow et al. 1990). When cultures approach saturation and the glucose in the media is used up, lactose uptake can occur and induction occurs more gently.

The use of AIM and lower temperatures provides three potential benefits to protein expression: lower temperature helps with protein folding, slower induction prevents shock to the cells, and growth to saturation allows theoretically higher yields of

protein per liter of culture. Indeed, when GST-M33 fusion protein was expressed using the AIM method, milligram quantities of protein were obtained. This solved one technical problem, however other problems remained. After elution from glutathione sepharose resin, there were quantitative amounts of degraded protein and significant nucleic acid contamination. To solve the issue of degradation, an additional affinity tag was included at the C-terminus of the open reading frame.

Numerous chemicals and reagents have been used to remove nucleic acid contaminations from extracts. These fall into two classes: nucleases and precipitation reagents. Since purified proteins would be used in assays containing nucleic acid, nucleases were excluded. Polyethylenimine (PEI) is a cationic polymer that has been used to precipitate nucleic acids from protein extracts (Burgess 2009). Low concentrations (0.15%) of PEI added to extracts efficiently precipitated nucleic acids. In initial trials at low salt concentrations non-reversible precipitation of the recombinant protein occurred. This was overcome by increasing the salt concentration to 1.6 M.

Using these conditions I was able to routinely get approximately 90% pure protein at milligram quantities per liter. These techniques proved to be generally useful for the expression and purification of PcG proteins as the frog, fish, and worm proteins were also successfully expressed and purified using these techniques.

Future directions

The analyses presented here demonstrate that chromatin compaction and inhibition of remodeling activities by PRC1 proteins have been conserved over at least five different organisms. These activities appear to depend on basic regions of the proteins that are predicted to be intrinsically disordered. Several experimental

methodologies can be taken to attempt to address the questions that remain: 1) What is the in vivo relevance of the charged domains of M33? 2) Do the intrinsically disordered regions of M33 become folded upon interacting with chromatin? 3) What are the precise molecular interactions between M33 and chromatin?

The studies presented in chapters two and three utilized in vitro assays to dissect the function of a mouse PcG protein. It will be important to follow up the observations using in vivo or cell culture systems. Since genetic knock-in experiments in murine systems are tedious, cell culture model systems should be a useful tool to determine whether the charged domain of M33 has biological relevance. Rescue experiments with wild type or mutant constructs followed with target gene analysis should allow the determination of biological relevance.

It is tempting to speculate that the apparently intrinsically disordered regions become ordered upon binding to chromatin, yet it is also possible that these domains would become ordered through interactions with other proteins in the PRC1 complex. PRC1 is on the order of megadaltons, so it remains possible that regions of M33 that we find important for observed in vitro activity are typically masked by other proteins. However, the similar results obtained from M33 alone or within in the core complex suggests that is not the case. Regardless, these issues warrant further consideration.

There are a number of different methods useful for characterizing intrinsically disordered proteins. Small-angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR) and deuterium exchange followed by mass spectrometry (DXMS) have all been used to this end (Lippens et al. 2006; Kathuria et al. 2011; Keppel et al. 2011). SAXS and NMR both require relatively concentrated protein samples and NMR has a functional

limit on the size of the protein being analyzed. DXMS provides a means to assay the relative accessibility of the peptide backbone, requires only minimal amounts of protein, and can give information about regions of polypeptides that become less accessible (ordered) upon binding to substrate. Any of these techniques should provide valuable information regarding the inherent flexibility of PcG proteins, and with DXMS it should be possible to map complex protein-protein and protein-DNA interaction motifs. These techniques provide intermediate details regarding the biophysical characteristics of proteins in contrast with the low resolution of the experiments described in chapters two and three and the high resolution afforded by X-ray crystallography.

Perhaps the biggest challenge with the highest reward will be high-resolution structural determinations of PcG proteins bound to repressed chromatin. Structural determination of PcG repressed chromatin—or of full length PcG proteins themselves—would provide insight into the molecular mechanisms at work during PcG mediated silencing. It could potentially demonstrate whether disordered domains in fact become folded upon interactions with substrate and whether basic patches make contacts with DNA or histones. A beautiful crystal structure of the BAH domain of Sir3 bound to a nucleosome provides evidence that these macromolecular structures can be determined, and that disordered loops can become ordered upon binding to nucleosomes (Armache et al. 2011). Recent work with M33 variants suggests that stable complexes can be formed with nucleosomes. I am hopeful that illuminating results will be forthcoming.

Appendix A: Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge

This paper was originally published in *Genes and Development* October 15, 2011.

Grau DJ, Chapman BA, Garlick JD, Borowsky M, Francis NJ, Kingston RE. 2011.
Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. *Genes Dev* **25**: 2210-21.

Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge

Daniel J. Grau,^{1,2} Brad A. Chapman,^{1,2} Joe D. Garlick,^{1,2} Mark Borowsky,^{1,2} Nicole J. Francis,³ and Robert E. Kingston^{1,2,4}

¹Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; ²Department of Genetics, Harvard Medical School, Harvard University, Cambridge, Massachusetts 02138, USA; ³Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

Polycomb group (PcG) proteins are required for the epigenetic maintenance of developmental genes in a silent state. Proteins in the Polycomb-repressive complex 1 (PRC1) class of the PcG are conserved from flies to humans and inhibit transcription. One hypothesis for PRC1 mechanism is that it compacts chromatin, based in part on electron microscopy experiments demonstrating that *Drosophila* PRC1 compacts nucleosomal arrays. We show that this function is conserved between *Drosophila* and mouse PRC1 complexes and requires a region with an overrepresentation of basic amino acids. While the active region is found in the Posterior Sex Combs (PSC) subunit in *Drosophila*, it is unexpectedly found in a different PRC1 subunit, a Polycomb homolog called M33, in mice. We provide experimental support for the general importance of a charged region by predicting the compacting capability of PcG proteins from species other than *Drosophila* and mice and by testing several of these proteins using solution assays and microscopy. We infer that the ability of PcG proteins to compact chromatin in vitro can be predicted by the presence of domains of high positive charge and that PRC1 components from a variety of species conserve this highly charged region. This supports the hypothesis that compaction is a key aspect of PcG function.

[Keywords: Polycomb; chromatin; transcriptional repression; intrinsic disorder]

Supplemental material is available for this article.

Received June 23, 2011; revised version accepted September 12, 2011.

The coordinated regulation of development requires the faithful maintenance of gene expression programs through multiple cell divisions. Two classes of proteins that maintain the epigenetic inheritance of gene states are the Polycomb group (PcG) and Trithorax group (TrxG) proteins (Kennison 1995; Schuettengruber et al. 2007). These proteins act in complexes that either repress or activate gene transcription, respectively (Lewis 1978; Ingham and Whittle 1980; Kennison and Tamkun 1988; Kennison 1993). PcG protein complexes map to hundreds of genomic loci, the most notable being the *Hox* clusters (Boyer et al. 2006; Bracken et al. 2006; Lee et al. 2006; Schwartz et al. 2006). These proteins form several different complexes, of which the best characterized are Polycomb-repressive complex 1 (PRC1) and PRC2 (Shao et al. 1999; Czermin et al. 2002; Kuzmichev et al. 2002; Muller et al. 2002). In current models of PcG-mediated repression, PRC2 is recruited to target loci, where it methylates

Lys 27 of histone H3 (H3K27me3) (Cao and Zhang 2004). This histone modification acts as a binding site for the PRC1 protein Polycomb (PC), although there are indications that other as-yet-uncharacterized mechanisms are also involved in targeting PRC1 action (Muller and Verrijzer 2009; Simon and Kingston 2009; Morey and Helin 2010).

Binding of PRC1 to target loci is believed to be central to the establishment of transcriptional silencing that is stable through cell divisions. The mechanisms via which PRC1 establishes repression are an area of intense study. PRC1 was first defined in *Drosophila*, where genetic studies initially identified the PcG genes via the phenotype of extra sex combs on the hind legs of male flies (Slifer 1942; Lewis 1947). There are multiple PRC1 family complexes in mammals. Each has some combination of four proteins, encoded by genes homologous to the *Drosophila* genes *Pc*, *Ph*, *Psc*, and *Sce* (*dRing*) (Cao et al. 2005; Schwartz and Pirrotta 2008). Potential mechanisms by which PRC1 family complexes silence genes include ubiquitylating histone H2A to initiate a block to transcription, directly inhibiting the transcriptional machinery, and creating a compacted state in chromatin that is

⁴Corresponding author.

E-mail kingston@molbio.mgh.harvard.edu.

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.172882.11>.

refractory to transcription. Here we explore chromatin compaction by PRC1 and the domain within PRC1 responsible for compaction.

In this study, compaction is defined as the ability to reduce the average distance of nucleosomes from each other as compared with the typical “beads on a string” seen with electron microscopy (EM). In vitro, compaction is thought to be directly related to the ability of PRC1 to inhibit chromatin remodeling, as the ability of PRC1 proteins to perform these two functions is highly correlated. This in vitro work extends to in vivo observations, since the ability of the *Drosophila* Posterior Sex Combs (PSC) protein to compact nucleosomal arrays correlates with the phenotypes of a set of mutations in PSC (King et al. 2005). In other work, PRC1 is suggested to stabilize nucleosomal turnover rates and create compacted chromatin domains large enough to be detectable by light microscopy in cells (Deal et al. 2010; Eskeland et al. 2010). These studies indicate that compaction is likely to be a biologically relevant mechanism of silencing by PRC1 family complexes.

If compaction is central to PRC1 function, then the ability to compact nucleosomal arrays should be conserved across organisms that contain PRC1. Here we show that the protein that is primarily responsible for compaction in mouse PRC1 is M33 (Cbx2), a homolog of *Drosophila* PC (Pearce et al. 1992). This was surprising, as PC is not a homolog of PSC or Su(z)2, the proteins responsible for compaction in *Drosophila* PRC1 (Francis et al. 2004; Lo et al. 2009). We performed a structure/function analysis of M33 and found that it and *Drosophila* PSC share a region that is highly basic and predicted to have a disordered secondary structure. Using protein charge as a basis, we identified putative PRC1 components in other organisms that are expected to compact nucleosomes and showed that these are functional in both inhibition of remodeling and compaction. These studies define a region in PRC1 proteins that functions similarly to the *Drosophila* protein PSC. We provide evidence supporting the idea that during evolution this key aspect of PRC1 function diverged onto distinct subunits. That this region appears to be present across evolution is consistent with it playing a key role in PRC1 function.

Results

M33 is a functional homolog of PSC

The *Drosophila* PcG protein PSC is able to block remodeling and compact nucleosomes in vitro, activities that might directly contribute to PRC1-mediated repression (King et al. 2002; Francis et al. 2004). We hypothesized that if these activities are important to PRC1 function, they would be conserved in mammalian PRC1. To investigate this, we used the mouse PRC1 core complex (mPCC). As in *Drosophila*, a PRC1 core complex of M33, Ring1A, and Bmi1 retains the majority of activity, so we chose to focus on these three subunits (Francis et al. 2001; Lavigne et al. 2004). We began by expressing mPCC and individual subunits of the core complex (Fig. 1A) to

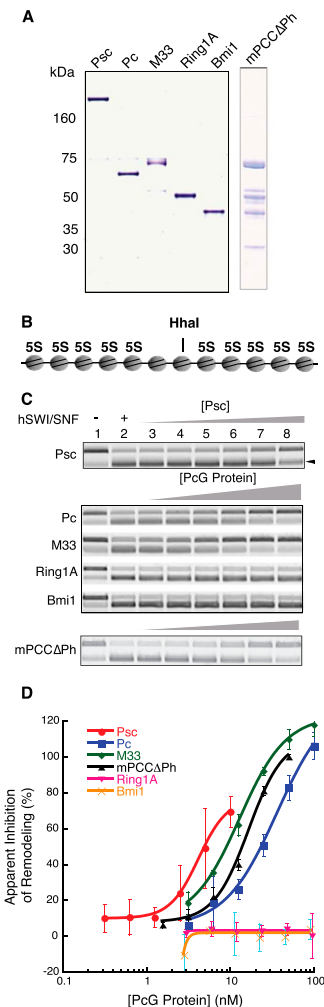


Figure 1. Identification of M33 as a functional homolog of PSC. (A) Coomassie-stained gel of PcG proteins purified from overexpression in *Sf9* cells. (B) Schematic representation of G5E4 nucleosomal array used in assays. (5S) 5S nucleosomal positioning sequence; (HhaI) unique HhaI restriction sequence that is inaccessible when packaged around a histone octamer. (C) Agarose gel of RIA assay. After reactions were completed, uncut and cut Cy5 end-labeled G5E4 DNA was separated on an agarose gel, scanned using a Typhoon PhosphorImager and quantified using ImageQuant software (GE Healthcare). The slower-migrating band represents the DNA that was not cleaved by HhaI (uncut), while the faster-migrating band represents cut DNA (arrowhead indicates cut DNA). (D) Graph of the data obtained by quantification of DNA bands in C. Graphs were created in Kaleidagraph software (Synergy) using a nonlinear sigmoidal curve fit. Error bars represent the standard deviation of three technical replicates. Apparent inhibition of remodeling was calculated by the following equation:

$$\frac{(\% \text{uncut with PcG and hSWI/SNF} - \% \text{uncut with hSWI/SNF})}{(\% \text{uncut without hSWI/SNF} - \% \text{uncut with hSWI/SNF})} \times 100.$$

determine which, if any, individual subunits had in vitro activity versus the core.

We characterized the activity of mPCC subunits and subcomplexes using a solution assay that we previously used to characterize PcG proteins. The restriction en-

zyme accessibility (REA) assay measures the ability of PcG proteins to antagonize nucleosome remodeling by the ATP-dependent remodeling complex hSWI/SNF (Kwon et al. 1994; Francis et al. 2001). Briefly, we assembled nucleosomes onto a 2.5-kb DNA/chromatin template (G5E4) (Fig. 1B) using salt dialysis (Uteley et al. 1998). This DNA fragment contains 10 5S nucleosomal positioning sequences. At the center of this fragment is space for two additional nucleosomes, for a total of 12. With two nucleosomes positioned in the central part of G5E4, one of them occludes a unique HhaI restriction site and prevents cutting. Movement of this nucleosome, which is accomplished efficiently by ATP-dependent nucleosome remodeling, allows HhaI to cut the G5E4 DNA. Preincubation of the arrays with PcG proteins prevents efficient remodeling of nucleosomes by the hSWI/SNF remodeling complex and therefore inhibits cutting by HhaI. By titrating in PRC1 proteins or complexes and measuring the amount of cut and uncut DNA, we are able to quantify their inhibitory activity.

As anticipated from previous studies, when we preincubate nucleosomal arrays with PSC or with a mouse core complex composed of Bmi1, Ring1A, and M33 (mPCC), we see a concentration-dependent inhibition of remodeling (Fig. 1C,D). The inhibition of remodeling activity that we observe with mPCC was lower than what we previously measured, which could reflect differences in the activities of protein preparations (Lavigne et al. 2004).

To determine the subunits of mouse PRC1 responsible for inhibition of remodeling activity, we tested proteins individually using this protocol. Surprisingly, we did not observe any appreciable inhibition of remodeling by Bmi1, the mouse homolog of the biochemically active PSC. Likewise, Ring1A did not exhibit any activity. However, M33, the mouse homolog of *Drosophila* PC, exhibited activity that was similar in efficiency to the core complex and to PSC (Fig. 1C,D). We conclude that the most active subunit in mouse PCC for inhibition of remodeling is the M33 protein.

Inhibition seems to require the presence of nucleosomes, as preincubation of M33 with naked DNA template prior to adding HhaI results in inhibition of cleavage that is two orders of magnitude less than when the DNA template is assembled into nucleosomes, suggesting that M33 is not directly interfering with the restriction enzyme (Supplemental Fig. S1A,B).

Inhibition of remodeling reactions was performed at ionic concentrations of 60 mM, which is optimal for hSWI/SNF activity (Supplemental Fig. S1C,D). We do not observe significant differences in inhibition of remodeling activity by M33 up to ionic conditions of 175 mM (Supplemental Fig. S1E,F).

These results led us to question the conservation of domain structure between *Drosophila* PRC1 and mammalian PRC1. The *Drosophila* PC protein can inhibit remodeling of nucleosomal templates, but is at least fivefold less efficient than PSC, while the mouse homolog of PC (M33) is at least one order of magnitude better at inhibiting remodeling than the mouse homolog of PSC

(Bmi1) (Fig. 1C,D) (note that Bmi1 displays a low level of compaction activity when assessed by EM; see below). We previously located the inhibition of remodeling and compaction activities in PSC to its C terminus, a region with no obvious primary sequence homology with any of the PcG proteins in the mammalian complex. We therefore set out to complete a structure/function analysis of M33 to determine what features of M33 were required for activity and how those features compared with PSC.

None of the conserved features of M33 are required for in vitro activity

To identify the domain of M33 required for inhibition of remodeling, we expressed and purified M33 and deletion variants in *Escherichia coli*. By using a cleavable GST tag on the N terminus and a Flag tag on the C terminus, we were able to obtain M33 and variants that were more homogeneous than the proteins we obtained using the baculovirus system (Supplemental Fig. S2A). We constructed a series of N-terminal and C-terminal truncation mutants, expressed and purified them, and tested their activity in the REA assay (Fig. 2A). In agreement with the above results, full-length M33 purified from *E. coli* has an inhibitory activity that is similar to M33 purified from Sf9 cells (cf. Figs. 1D and 2B).

M33 does not share significant sequence alignment with PSC, but contains several domains that are conserved in other chromatin-binding proteins: a chromodomain (CHD), an AT-Hook (ATH), and a PC C-box (CBOX) (Fig. 2A). Another domain, termed here as the homology domain (HD; amino acids 252–266), is a motif that is conserved in M33/Cbx2 homologs but is not found in other Cbx proteins.

We first determined whether any known domains in M33 contributed to its inhibitory activity. Deletion of either the ATH or HD motifs did not result in any reduction in inhibition activity (data not shown). The CHD and CBOX domains interact with nucleosomes, and additionally, the CBOX domain is required for repression activity in *Drosophila* embryos and in transient transfection assays, as well as for interactions with RING proteins (Muller 1995; Schoorlemmer et al. 1997; Breiling et al. 1999; Satijn and Otte 1999; Fischle et al. 2003; Min et al. 2003; Wang et al. 2008). Surprisingly, deletion of the CHD or CBOX domains did not abolish the inhibition activity (Fig. 2B,C; Supplemental Fig. S2B).

Since none of the conserved features of M33 are required for our in vitro activity, we wondered whether there might be a nonconserved motif that was responsible for the inhibition of remodeling activity. Thus, we decided to create a series of N-terminal and C-terminal deletions.

We cloned and expressed a series of truncation mutants of the M33 protein and tested them in the REA assay for inhibition activity (Fig. 2B,C; Supplemental Fig. S2B). We found that truncations of the C terminus to amino acid 105 (M33^{1–105}) still retain repression activity. In contrast, N-terminal deletions to amino acid 248 (M33^{249–519}),

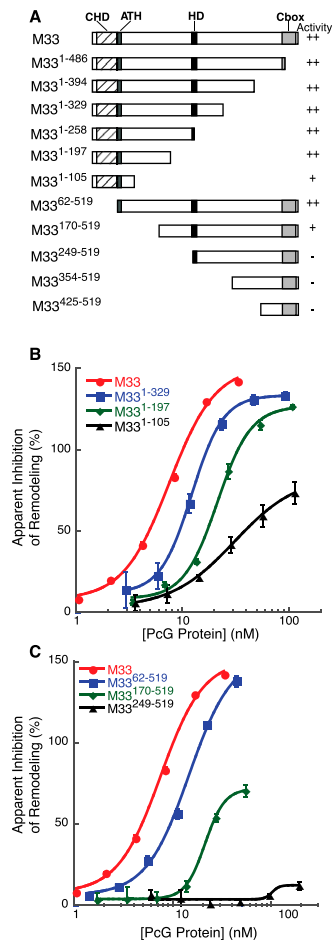


Figure 2. Structure/function analysis of M33-mediated repression. (A) Diagram of M33 truncation mutants tested for biochemical activity. (B) Graph of inhibition activity of selected M33 C-terminal truncation mutants. Data were analyzed as in Figure 1. (C) Graph of inhibition activity of selected M33 N-terminal truncation mutants.

completely abolishes activity. Since the M33¹⁻¹⁹⁷ and M33⁶²⁻⁵¹⁹ constructs retain near wild-type levels of activity, we conclude that an N-terminal region between amino acids 62 and 197 is required for optimal repression activity.

Positively charged residues are required for optimal activity in vitro

We wondered what characteristics are shared between the regions of M33 and PSC that are required for inhibition of remodeling. Examination of the amino acid composition of M33 revealed a bias toward lysines and arginines—the two most positively charged amino acids. The C-terminal region of PSC that is required for in vitro and in vivo activity also has an overrepresentation of arginines and lysines. The predicted overall charge of M33 and PSC is +32.5 and +82.1, respectively. This is in contrast to Bmil, inactive in the REA assay, which has a predicted overall charge of +10.1.

We wondered whether we could observe any correlation between the predicted overall charge of the M33 variants and inhibition of remodeling activity. Indeed, if we plot the predicted charge of the truncation mutants versus the concentration required for 50% inhibition of remodeling, we obtain a linear relationship with $R^2 = 0.7$ (Fig. 3A). These charged amino acids are spread throughout M33 and PSC, which is reminiscent of another class of repressive proteins: the linker histones (Hansen et al. 2006).

We hypothesized that these basic residues in M33 are important for binding to chromatin and therefore contribute to PcG-mediated repression. To test this hypothesis, we synthesized mutant variants of the M33 protein that perturb overall protein charge. We systematically mutated every other lysine or arginine in M33¹⁻⁴⁸⁶ to alanine, beginning with K132A and ending with K240A, termed here M33^{1-486,1KR_A}, for a total of 13 residues mutated (Fig. 3B; Supplemental Fig. S3B,C). We used the M33¹⁻⁴⁸⁶ construct for this purpose because we found that deletion of the C-terminal 33 amino acids of M33 enhances expression and purification with marginal effect on in vitro activity. Additionally, we generated a construct that mutates every other arginine or lysine from R262A to R415A (M33^{1-486,2KR_A}), as well as a combination of the two regions (M33^{1-486,1KR_A2KR_A}). These constructs reduce the predicted overall charge of M33¹⁻⁴⁸⁶ at pH 7.0 from +32.5 to +19.3, +22.3, and +9.3, respectively. Finally, we synthesized a construct in which every other aspartic or glutamic acid was mutated to alanine, from E156A to E399A, for a total of seven mutated residues (M33^{1-486,1DE_A2DE_A}). This construct is expected to increase the predicted charge to +39.3. If overall positive charge of these domains is important to function, these mutations are expected to reduce and enhance the in vitro activity of M33, respectively.

We expressed and purified these proteins to a level similar to M33¹⁻⁴⁸⁶ (Fig. 3C). When we tested these proteins for inhibition of remodeling activity, we found that M33^{1-486,1DE_A2DE_A} has inhibition activity similar to M33¹⁻⁴⁸⁶ (Fig. 3D; Supplemental Fig. S2C). Additionally, M33^{1-486,2KR_A} inhibits remodeling to an extent similar to M33¹⁻⁴⁸⁶. However, both M33^{1-486,1KR_A} and M33^{1-486,1KR_A2KR_A} reduce inhibition activity, approximately threefold in the case of M33^{1-486,1KR_A} and fivefold for M33^{1-486,1KR_A2KR_A} (data for all of the charge mutants are depicted as blue circles in Fig. 3A). These results are consistent with the idea that the charged residues within the M33^{1-486,1KR_A} region are important for optimal repression activity and agree with the deletion analysis above that suggests that the region between amino acids 62 and 197 is required for optimal repression activity. However, since there was no enhancement of activity in the M33^{1-486,1DE_A2DE_A} variant and no loss of activity in the M33^{1-486,2KR_A} mutant, we conclude that there must be criteria other than simple overall protein charge contributing to inhibition activity of M33.

Since PcG proteins typically function within complexes, we tested whether the results we obtained with

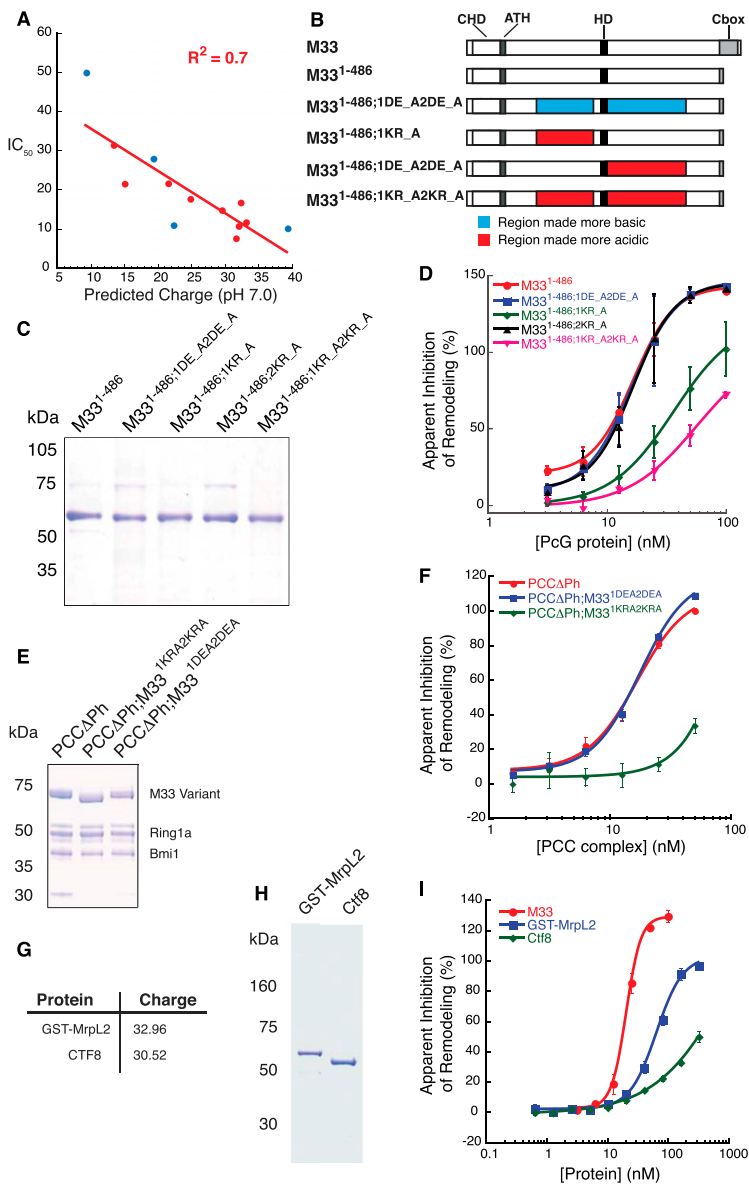


Figure 3. The role of charge in M33-mediated repression activity. (A) Graph of predicted protein charge at pH 7.0 versus IC_{50} as determined using Kaleidagraph software and performing a linear fit. Red circles represent data from M33 truncation mutants, and blue circles represent data from M33 charge mutants. (B) Schematic representation of the charge mutant proteins that were tested. (C) Coomassie-stained gel of M33 charge mutants expressed and purified from *E. coli*. (D) Plot of the quantification from the REA done with M33 charge mutants. (E) Coomassie-stained gel of M33 charge mutants in the context of the core PRC1 complex. Proteins were expressed and purified from *Sf9* cells. (F) Plot of the quantification from the REA done with the charge mutant complexes. (G) Charge characteristics of the non-PcG basic proteins cloned. The accession numbers for the proteins are MrpL2, NP_079578.1; and CTF8, AAH23107.1. (H) Coomassie-stained gel of the non-PcG basic proteins expressed and purified from *E. coli*. (I) Plot of the quantification from the REA done with the non-PcG basic proteins.

M33 alone could be replicated in the context of the core complex. We coexpressed two full-length M33 charge variants with Bmi1 and Ring1A in *Sf9* cells and were able to purify both mutant complexes to a level similar to the wild-type complex (Fig. 3E). In agreement with the results obtained with M33 variants alone, PCCΔPh;M33^{1DE_A2DE_A} inhibited remodeling to an extent similar to the wild-type core complex, whereas PCCΔPh;M33^{1KR_A2KR_A} reduces activity by more than fourfold (Fig. 3F; Supplemental Fig. S2C).

We wondered whether inhibition activity is a general characteristic of any basic protein. We cloned two mouse genes that contain charge characteristics similar to M33 and expressed the proteins MrpL2, a mitochondrial protein with predicted charge of +32.96, and CTF8, a nuclear protein with a predicted charge of +30.52. These non-PcG proteins are sevenfold to 10-fold less active than M33, suggesting that

features beyond overall charge are involved in PcG protein activity (Fig. 3G–I; Supplemental Fig. S2E).

To determine whether there are any conserved sequence motifs within the region mutated in the M33^{1-486;1KR_A} construct, we performed a sequence alignment with frog, zebrafish, and chicken Cbx2 proteins (Supplemental Fig. S3A). We found that five of the mutated residues are conserved among all four species, raising the possibility that these amino acids are involved in nucleosome interactions. While we cannot rule out that the loss of activity is due to disruption of protein structure, these results are consistent with the idea that the intrinsic charge of PcG proteins is important for repression activity.

Given the high number of hydrophilic amino acids in M33, one possibility is that the charged region is “natively unfolded” or “intrinsically disordered” (Uversky

and Dunker 2010). Natively unfolded regions occur in other chromatin architectural proteins, including the PcG proteins RYBP and GAGA factor, and have been proposed to play a role in the function of PSC and Su(z)2 (Agianian et al. 1999; Emmons et al. 2009; Lo et al. 2009; Neira et al. 2009). The linker histone regions required for chromatin compaction are intrinsically disordered in solution. Genome-wide predictions of natively unfolded regions in proteins forecast a high percentage of transcriptional regulators as having some intrinsic disorder (Garza et al. 2009; Sandhu 2009). Consistent with this idea, Metadisorder, a program that uses several different disorder prediction algorithms, predicts that M33 is folded within the CHD and CBOX, while the central portion that contains the 1KR_A region is disordered (Supplemental Fig. S4A; Kurowski and Bujnicki 2003). Additionally, the far-UV circular dichroism (CD) spectra of M33¹⁻⁴⁸⁶ demonstrate that it has the characteristic spectra of a protein that contains intrinsic disorder; specifically, a minima of molar ellipticity at 200 nM (Supplemental Fig. S4B).

Evolution of PcG proteins suggests charge is a predictor for *in vitro* activity

The above studies collectively suggest that a positively charged domain is responsible for the ability of PRC1 proteins to inhibit nucleosome remodeling. *In vitro*, M33 appears to inhibit remodeling enzymes in a manner that is dose-responsive to the overall charge of the protein, whereas PSC appears to be more complicated. Two truncation mutants of PSC that disrupt overall charge to variable degrees do not directly correlate with inhibition activity. PSC¹⁻⁹⁰⁹ contains half as much overall charge as PSC⁴⁵⁶⁻¹⁶⁰³, yet inhibits remodeling to a similar extent (Lo and Francis 2010).

Since M33 and PSC are not sequence homologs, we hypothesized that these proteins evolved to become functional homologs. This hypothesis predicts that organisms that are more closely related to *Drosophila* will have PSC homologs that are active. Likewise, it predicts that organisms more related to mice will have M33 homologs that are classified as active. We sought to test these predictions by using a computational approach.

Homologous M33/PC and Bmi1/PSC proteins were identified by searching the UniProt database for conserved CHD (IPR000953) and RING (IPR001841) InterPro signatures (Hunter et al. 2009; The UniProt Consortium 2011). Identified proteins from this search were filtered to PcG-like proteins using additional regions of homology—the CBOX domain for M33/PC and the extended RING HD for Bmi1/PSC—by using hmmsearch (Eddy 2009).

Using PSC and M33 as known active proteins and Bmi1 as a known inactive protein, the PcG proteins were classified as active or nonactive using k-means clustering (Cock et al. 2009). The clustering was based on overall protein charge and regional charge. Overall charge was calculated using Biopython, and regional charge was calculated as the percentage of 75-amino-acid windows with an isoelectric point (pI) >10.2. Proteins of the active

class contained a regional charge of at least 10% and an overall charge of at least +15. Additional classification parameters such as protein interaction and domain distribution were considered, but were excluded as noninformative.

The total number of CHD and RING domain proteins identified were 44 and 59, respectively, from 13 species (Supplemental Table S1). Of these proteins, 32 CHD proteins and four RING proteins are predicted to inhibit remodeling. We then used 18S ribosomal RNA (rRNA) sequences from these 13 species to generate a phylogenetic tree to show the evolutionary relationships of these organisms (Fig. 4A). Interestingly, when we did this, we observed an evolutionary point where the predicted

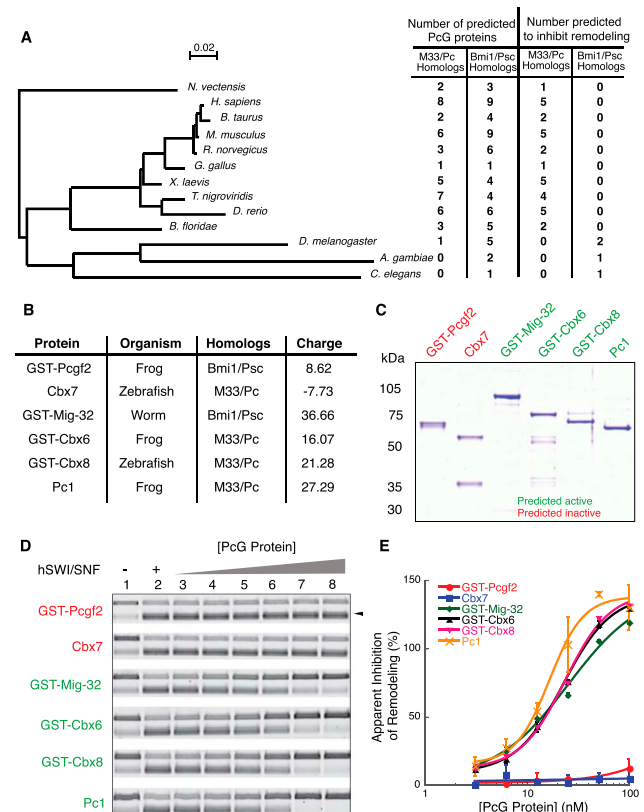


Figure 4. Analysis of evolutionary conservation of PcG function. (A) Phylogenetic tree of species containing RING domain or CHD proteins from UniProtKb protein database. The tree is based on alignments of 18S rRNA from each of the species. Number of predicted PcG proteins represents the number of each class of proteins that was found in the UniProtKb database. Number predicted to inhibit remodeling is the number of proteins from each class that is expected to have inhibition activity based on overall protein charge and regional charge. The bar represents 0.02 substitutions per site. (B) Charge properties of PcG proteins selected for *in vitro* activity analysis. The accession numbers for the proteins are Pcgf2, NP_001084738.1; Cbx7, NP_001017853.1; Mig-32, NP_502293.2; Cbx6, NP_001088074.1; Cbx8, AAI54356.1; and Pc1, NP_001081900.1. (C) Coomassie-stained gel of PcG proteins expressed and purified from *E. coli*. (D) Agarose gel of REA assay reaction products. (E) Plot of the quantification of results obtained in D.

activity of RING-containing PcG proteins and CHD-containing proteins diverged. In deuterostomes, the PcG proteins predicted to be active were all CHD-containing. In protostomes, only predicted active RING domain-containing proteins were observed. Notably, despite the expansion of RING proteins in deuterostomes, we could not observe any Bmi1/PSC homologs that the analysis predicted to be active. We used this list to test the accuracy of our predictions.

To accomplish this, we expressed and purified six evolutionarily divergent CHD- and RING-containing PcG proteins and tested their activity in vitro. These proteins were from the species *Xenopus laevis*, *Danio rerio*, and *Caenorhabditis elegans* (Fig. 4B,C). Overall, we tested one predicted inactive RING protein (frog Pcgf2), one predicted inactive Cbx protein (zebrafish Cbx7), one predicted active RING protein (worm Mig-32), and three predicted active Cbx proteins (zebrafish Cbx8, frog Pc1, and Cbx6) (see Fig. 4B for predicted charge).

Proteins were chosen based on evolutionary divergence from either mice or *Drosophila*, and the availability of cDNAs. When we tested the activity of these proteins using the REA assay, we found that, as predicted, frog Pcgf2 and zebrafish Cbx7 do not inhibit remodeling activity (Fig. 4D,E). Conversely, preincubation with arrays using worm Mig-32, frog Pc1 or Cbx6, or zebrafish Cbx8 leads to efficient inhibition of remodeling. We did not observe any effect from leaving the GST tag on our proteins or for GST alone (Supplemental Fig. S5). Thus, for these proteins, regional charge was an accurate predictor of in vitro activity.

Compaction of chromatin by divergent PcG proteins

Our results above are consistent with the hypothesis that the domain primarily responsible for inhibition of remodeling in vitro resides on the PSC homolog in flies and worms and resides on the PC homolog in vertebrates. This idea posits that one subunit of PRC1—either an M33/PC homolog or a Bmi1/PSC homolog—will contain a region (or regions) of high positive charge that is important to repression of remodeling and also for compaction. To test whether these proteins are able to compact nucleosomal arrays, we used EM to visualize the interaction of PcG proteins with arrays.

PSC can efficiently compact nucleosome arrays, as observed visually by EM and as quantified by measuring the end-to-end distance of arrays in the EM images (Francis et al. 2004). We used this technique to examine M33 and the evolutionarily conserved PcG proteins to see which of these proteins could compact chromatin in a manner like PSC. EM images of PcG proteins preincubated with G5E4 arrays were acquired, and images of low quality were discarded in a double-blind manner. Nucleosomal arrays alone look like the canonical “beads on a string” conformation (Fig. 5A). Preincubation with Bmi1 tends to reduce the overall array length, but does not appear to promote intranucleosomal interactions. M33, in contrast, promotes the formation of compact particles consisting of multiple nucleosomes.

To quantify the ability of these proteins to compact arrays, a single-blind measurement of the end-to-end length of protein/array particles was performed. Full-length M33 creates compacted nucleosome structures as determined by a significant decrease in end-to-end length when compared with arrays incubated with no protein (Student's *t*-test, *P*-value < 0.0001) (Fig. 5B). Surprisingly, despite being inactive in the REA, Bmi1 is also able to measurably compact arrays. However, M33 is able to compact chromatin significantly better than Bmi1 (*P*-value < 0.0001).

Next we quantified the ability of the PcG proteins from divergent species to compact arrays. In agreement with the previous REA assay, neither GST-Pcgf2 nor Cbx7 (predicted inactive) is able to significantly compact arrays (*P*-values = 0.1 and 0.8, respectively) (Fig. 6B). These reactions yield arrays that have extended conformations. However, the proteins that inhibit remodeling (GST-Mig32, GST-Cbx6, GST-Cbx8, and Pc1) are able to significantly compact arrays (*P*-value < 0.0001) (Fig. 6B). As with M33, these proteins promote intranucleosomal interactions.

Thus, we predicted PcG protein activity based on regional charge and showed that the predictions appear to hold true. Proteins that we predict to be active both inhibit remodeling and compact nucleosomes in vitro. In contrast, the predicted inactive proteins do not.

Discussion

Here we show that the predicted protein charge of a mouse PcG protein correlates with in vitro activity.

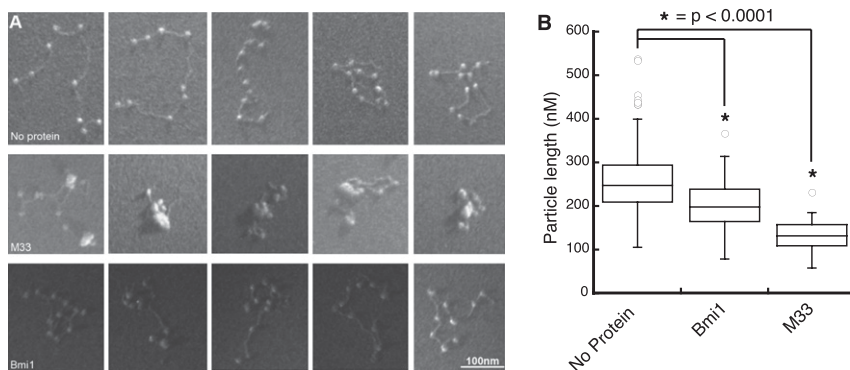


Figure 5. Compaction of nucleosomal arrays by mouse PcG proteins. (A) Representative EM images of nucleosomal arrays incubated with the indicated PcG protein. (B) Box plot representation of the measured maximal diameter of nucleosomal array particles. Particle length is the diameter of the smallest circle that can entirely surround one nucleosomal array. The box represents the upper and lower quartile, and the line splitting the box represents the mode. The open circles represent outliers, and the asterisks indicate a *P*-value of <0.0001 using Student's *t*-test. No protein, *n* = 72; Bmi1, *n* = 50; M33, *n* = 30.

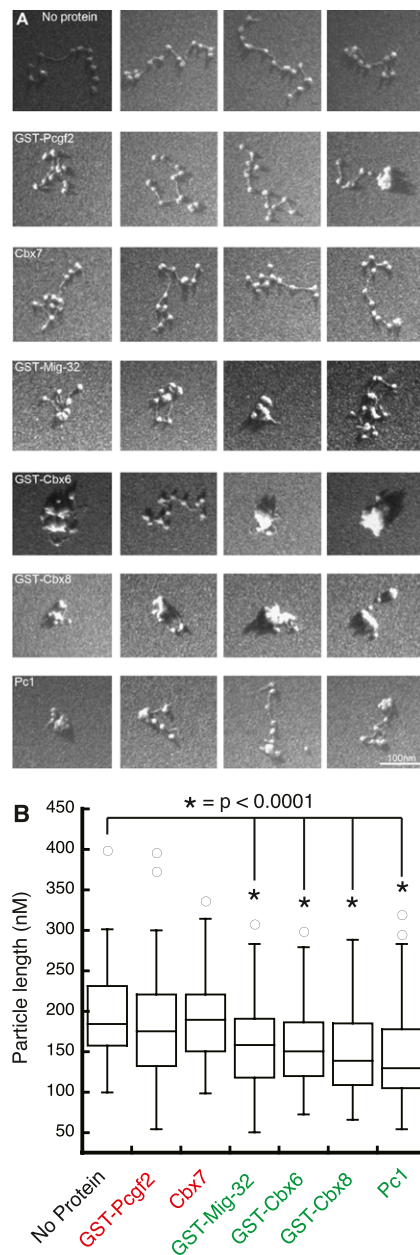


Figure 6. Compaction of nucleosomal arrays by diverse PcG proteins. (A) Representative images of nucleosomal particles incubated with various PcG proteins from different species. (B) Box plots of images as described above. No protein, $n = 79$; GST-Pcgf2, $n = 86$; Cbx7, $n = 113$; GST-Mig-32, $n = 90$; GST-Cbx6, $n = 88$; GST-Cbx8, $n = 87$; Pc1, $n = 89$.

We extended this observation by making a computational prediction of PcG activity in a variety of species and demonstrated that we can predict activity based on charge characteristics. These results support the hypothesis that one key function for PRC1 proteins is the ability to compact nucleosomal arrays and repress chromatin remodeling. The conservation of this basic, charged domain suggests that it may be important to silencing by PRC1 family proteins.

Roles of natively unfolded proteins

Natively unfolded or intrinsically disordered proteins were first described in the late 1980s (Sigler 1988). These early descriptions were focused on the proteins that are involved in transcriptional activation. Notably, it was observed that the negatively charged amino acids of proteins required for optimal transcriptional activation did not need to be precisely ordered; Ma and Ptashne (1987) elegantly demonstrated that the critical parameter appeared to be amino acid composition. We found that canonical transcription repressors, the PcG proteins, also appear to have regions of disorder, yet, in contrast to transcriptional activators, contain high concentrations of basic amino acids. It is tempting to speculate that these oppositely charged disordered regions play a “yin-yang” role in transcriptional regulation. It is possible that, in addition to the roles in nucleosome interaction described above, these positively charged transcription repressors could directly interact with and inhibit the negatively charged activation domains of the transcriptional machinery.

There are several proposed reasons why proteins would contain regions of disorder. Disordered regions could potentially adopt different conformations that allow interactions with multiple binding partners. This “hub” function is expected to be beneficial for regulatory proteins; a single protein could potentially regulate many different proteins in a context-specific manner (Gunasekaran et al. 2003). There is also the “fly casting” model, where an extended conformation could allow a protein to “sample” a larger amount of space, forming and breaking low-affinity contacts until conformational change induces tighter binding (Shoemaker et al. 2000). This is expected to promote interactions of low affinity and high specificity. One computational predictor of protein disorder—charge—was found to also be predictive of PcG functional activity, suggesting that charged disordered regions could possibly play a general role in PcG-mediated repression.

Charged domains and PcG function

What might be the biological role for PcG charged domains in the repression of transcription? They appear to be predictive for both inhibition of remodeling and compaction of chromatin in vitro. Here we propose a model for how the charged domains of PRC1 function: (1) PRC1 is recruited to target loci and presents the charged domain to linker and/or nucleosomal DNA (Fig. 7A). (2) The charged domain initially interacts with a nucleosome and creates more interactions with other nucleosomes (Fig. 7B). (3) Finally, oligomerization occurs through Ph or other protein–protein interactions to promote spreading or formation of higher-order chromatin fibers (Fig. 7C).

The CBOX domain of M33 is not required for in vitro repression activities, yet this motif is conserved and required for the repression of template DNA in cell-based assays (Schoorlemmer et al. 1997). This domain is required for interactions with Ring1A/B and Bmi1, which in turn interact with Ph proteins (Alkema et al. 1997;

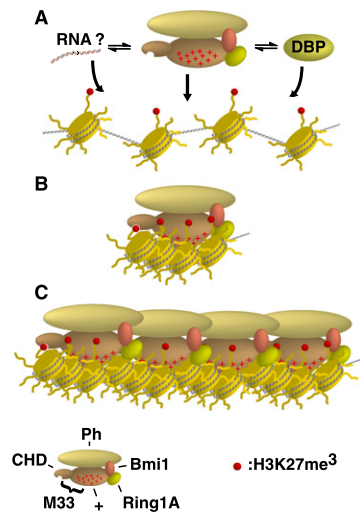


Figure 7. A model for chromatin compaction by the mPCC. (A) The mPCC is recruited to target loci, potentially through a variety of mechanisms. (DBP) DNA-binding protein. (B) The charged region of M33 (indicated by plus signs) interacts with nucleosomes to compact chromatin. (C) Further protein–protein interactions from other proteins in the core PRC1 complex drive spreading of compacted chromatin.

Gunster et al. 1997; Hashimoto et al. 1998; Hemenway et al. 1998; Satiijn and Otte 1999). Thus, we imagine that an initially transient nucleosome–nucleosome interaction mediated by charged domains facilitates the further stabilization of a repressed chromatin structure that is mediated by other PRC1 proteins. Studies analyzing the dynamics of Cbx/chromatin interactions in culture cell models observe both transiently and stably associated Cbx proteins, consistent with an initial unstable interaction followed by step(s) that promote stable associations (Ren et al. 2008).

PcG protein evolution

What might be the explanation for how the charged domain evolved to reside on an M33/PC homolog in mammals as opposed to a Bmi1/PSC homolog in flies? We can imagine at least four possible explanations: (1) An early common ancestor had both an M33/PC and a Bmi1/PSC homolog that each contained a charged region; sometime during evolution, the characteristic charge was lost from one or the other homolog. (2) The charged region initially resided on either an M33/PC homolog or a Bmi1/PSC homolog in the common ancestor, but was lost during evolution and subsequently gained on the other PcG homolog. (3) The charged region initially resided on either an M33/PC homolog or a Bmi1/PSC homolog in the common ancestor and was gained on the other before being lost. (4) A convergent evolutionary event occurred: Neither the M33/PC nor the Bmi1/PSC homolog in the common ancestor had a charged domain, but during evolution, one homolog or the other evolved it.

Resolution of the possibilities mentioned above will require a more extensive examination of potential PcG

members across evolution than has been performed here. Our phylogenetic tree was not a comprehensive list of all deuterostomes and protostomes, so it is possible that there may be unexpected active homologs in other species. Additionally, protein sequences may have been incorrectly annotated in the database that we queried or missing from organisms with incomplete genomes. One unclear example involves Su(z)2, a homolog of PSC that can inhibit remodeling *in vitro*. Su(z)2 was found in both the predicted active and inactive lists, although the isoforms in the inactive list appear to be short fragments of the full-length protein that may not exist *in vivo*. Regardless of a full examination of this issue, the data reported here are consistent with the hypothesis that there was evolutionary pressure to maintain a highly charged domain in a component of PRC1. One possibility is that across multiple species, the core of PRC1 will contain several conserved domains/motifs—two RING fingers, a CHD, a SAM domain, and a disordered/charged domain—defining the basic functional unit of PRC1.

Molecular nature of PcG/chromatin interactions

We do not understand the precise molecular mechanisms behind PcG protein interactions with chromatin. The flexible charged domains might interact with linker DNA, nucleosomal DNA, the histones themselves, or a combination of these chromatin components. We found that two non-PcG proteins with predicted charges similar to M33 do not inhibit remodeling as well as M33. This suggests a mechanism that does not rely solely on the amount of positive charge. It is possible that function involves a specific spacing of the charged residues and/or juxtaposition of the charged surface with other functional domains. For example, the majority of the active proteins that we characterized also contain a CHD, a known histone-binding domain, opening up the possibility that both DNA and histone contacts are required for optimal PcG-repressive activities.

Previous studies have shown that intrinsically disordered regions of proteins can become folded upon interacting with their substrate. This gives us hope that eventually the molecular mechanisms of chromatin condensation by PcG proteins can be unraveled using structural approaches.

Materials and methods

Protein expression

Baculovirus expression was done essentially as described (Francis et al. 2001). Briefly, Sf9 cells were grown in Hyclone CCMIII medium at 27°C with shaking. One liter of cells was either infected with individual viruses for the expression of single subunits or coinfecting with multiple viruses for the expression of protein complexes. After 40 h, cells were harvested by centrifugation and washed in PBS prior to making nuclear extracts (NEs) as described (Sif et al. 1998). For the purification of single PcG subunits, NEs were bound to M2 resin (Sigma, catalog no. A2220) for 4 h, then washed extensively with BC buffer (20 mM HEPES at pH 7.9, 0.2 mM EDTA, 20% glycerol, 1 mM DTT, 10 mM PMSF, protease inhibitor cocktail [Roche

complete protease inhibitor tablets)) containing 500 mM KCl. The M2 beads were then washed with BC buffer containing higher concentrations of KCl, up to 2 M, before eluting in BC buffer containing 500 mM KCl and 0.4 mg/mL Flag peptide. Purification of PcG complexes was identical, except that the bound protein was washed with BC buffer containing 300 mM KCl.

For expression of GST fusion proteins in *E. coli*, Rosetta pLysS cells were transformed with pGEX6P1 containing the cDNA of interest. A single colony was used to inoculate 5 mL of LB containing 25 µg/mL chloramphenicol and 50 µg/mL ampicillin and grown overnight. The following morning, 250 µL of the overnight culture was used to inoculate 250 mL of the auto-induction medium ZYP-5052 [12 g/L Bacto tryptone, 24 g/L Bacto yeast extract, 25 mM (NH₄)₂SO₄, 50 mM KH₂PO₄, 50 mM Na₂HPO₄, 0.05% glucose, 0.2% α-lactose, 0.5% glycerol, 1 mM MgSO₄, 25 µg/mL chloramphenicol, 50 µg/mL ampicillin] (Studier 2005). Cultures were shaken for 5 h at 37°C, then grown overnight at 18°C. The cultures were collected by centrifugation at 4000 rpm in a Beckman J6 MI for 20 min. Cell pellets were resuspended in 40 mL of lysis buffer (50 mM HEPES at pH 7.5, 0.5 mM EDTA, 1.6 M KCl, 20% glycerol, 0.5 mM MgCl₂, 0.05% NP40, 1 mg/mL lysozyme, 1 mM DTT, protease inhibitors). The cells were taken through three freeze-thaw cycles, then sonicated to shear DNA before centrifugation at 25,000g for 20 min to remove debris. Five percent polyethylenimine (PEI) in 20 mM HEPES (pH 7.5) was added dropwise to the supernatant while stirring to a final concentration of 0.15%, and stirred an additional 30 min. This step was omitted for GST-MrpL2. The precipitated nucleic acid was removed by centrifugation at 25,000g for 20 min. Extracts were bound to glutathione sepharose beads for 2 h before washing with BC buffer containing increasing salt as described above for M2 purifications. Proteins were either cleaved or eluted off of the resin. For cleavage of proteins, the resin was incubated overnight with 20 U of HRV 3C protease in 5 mL of BC buffer with 500 mM KCl. Proteins were eluted from glutathione sepharose by incubating in BC buffer containing 500 mM KCl and 40 mM reduced glutathione. Eluted proteins were purified over M2 resin as described above. Purified proteins were quantified using the Bradford assay, and then normalized relative to each other by the intensity of protein bands on an SDS-PAGE gel analyzed using ImageJ software.

REA assays

Nucleosome arrays were assembled using HeLa histones as previously described, except that Cy5-labeled G5E4 was used (Sif et al. 2001). Human SWI/SNF was purified from HeLa NEs as described (Sif et al. 1998). Reactions were carried out in a 20-µL volume containing 12 mM HEPES (pH 7.9), 12% glycerol, 60 mM KCl, 0.12 mM EDTA, 0.12 µg/µL BSA, 2 mM ATP, 1.25 mM MgCl₂, and 2 nM assembled nucleosomes. Dilutions of PcG proteins were incubated with the arrays for 30 min at 30°C prior to the addition of 100 ng of SWI/SNF and 8 U of HhaI. Reactions were incubated for 1 h at 30°C before the addition of 10 µL of stop buffer (1.5 mg/mL Proteinase K, 70 mM EDTA, 10 mM Tris at pH 7.7, 1% SDS, 0.1% orange G). Reactions were incubated for 30 min at 55°C, then separated on a 1% agarose gel in 1× TAE buffer. DNA was visualized on a Typhoon scanner and quantified using ImageQuant software. Apparent inhibition of remodeling was determined by the equation

$$\frac{(\% \text{uncut with PcG and hSWI/SNF} - \% \text{uncut with hSWI/SNF})}{(\% \text{uncut without hSWI/SNF} - \% \text{uncut with hSWI/SNF})} \times 100$$

and plotted using Kaleidagraph. Curves were generated using an equation for a sigmoidal fit.

Computational classification of PcG proteins

The UniprotKb database was queried for proteins containing CHDs (IPR000953) or RING domains (IPR001841). This list was filtered for PcG proteins using HMMER by keeping only those proteins containing a CBOX for CHD proteins or extended homology for RING proteins. The predicted charge and regional charge of known active and inactive PcG proteins (regional charge is defined as the percentage of sliding windows of 75 amino acids where the average pI is 10.2 or above) was used to cluster the unknown PcG proteins into either active or inactive classes using the k-means method. Proteins of the active class contained at least 10% regional charge and overall charge of 15 or greater. The reproducible Python scripts for this analysis are available (https://github.com/chapmanb/mgh_projects/tree/master/dg_PSC).

The phylogenetic tree was generated by using ClustalW to create an alignment of 18S rRNAs, and then using the maximum likelihood method. The tree was drawn using NJplot.

EM

EM experiments were carried out essentially as described previously (Francis et al. 2004), with the following minor modifications. Binding reactions were carried out in either 30 or 60 mM KCl, and NP40 added to a final concentration of 0.001%. All reactions were set up at molar ratios of eight PcG proteins to one nucleosome, based on total protein concentration. This ratio was selected based on electrophoretic mobility shift assays carried out under the same conditions as EM, which demonstrated binding of M33 to nucleosomal arrays with minimal aggregation at this ratio.

Acknowledgments

We thank Sarah K. Bowman, José M. Antao, Karim Bouazoune, and Leslie Beh for critically reading the manuscript and thoughtful discussions, and Anders Näär and C. Ting Wu for critical discussions and experimental suggestions. This work was supported by grants from the NIH (GM043901 to R.E.K., and GM078456-01 to N.J.F.).

References

- Agianian B, Leonard K, Bonte E, Van der Zandt H, Becker PB, Tucker PA. 1999. The glutamine-rich domain of the *Drosophila* GAGA factor is necessary for amyloid fibre formation in vitro, but not for chromatin remodelling. *J Mol Biol* **285**: 527–544.
- Alkema MJ, Bronk M, Verhoeven E, Otte A, van 't Veer LJ, Berns A, van Lohuizen M. 1997. Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. *Genes Dev* **11**: 226–240.
- Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**: 349–353.
- Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**: 1123–1136.
- Breiling A, Bonte E, Ferrari S, Becker PB, Paro R. 1999. The *Drosophila* polycomb protein interacts with nucleosomal core particles in vitro via its repression domain. *Mol Cell Biol* **19**: 8451–8460.
- Cao R, Zhang Y. 2004. The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Curr Opin Genet Dev* **14**: 155–164.

- Cao R, Tsukada YI, Zhang Y. 2005. Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol Cell* **20**: 845–854.
- Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422–1423.
- Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. 2002. *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**: 185–196.
- Deal RB, Henikoff JG, Henikoff S. 2010. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**: 1161–1164.
- Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. *Genome Inform* **23**: 205–211.
- Emmons RB, Genetti H, Filandrinos S, Lokere J, Wu CT. 2009. Molecular genetic analysis of Suppressor 2 of zeste identifies key functional domains. *Genetics* **182**: 999–1013.
- Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultschi AI, Wutz A, et al. 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell* **38**: 452–464.
- Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**: 1870–1881.
- Francis NJ, Saurin AJ, Shao Z, Kingston RE. 2001. Reconstitution of a functional core polycomb repressive complex. *Mol Cell* **8**: 545–556.
- Francis NJ, Kingston RE, Woodcock CL. 2004. Chromatin compaction by a polycomb group protein complex. *Science* **306**: 1574–1577.
- Garza AS, Ahmad N, Kumar R. 2009. Role of intrinsically disordered protein regions/domains in transcriptional regulation. *Life Sci* **84**: 189–193.
- Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, Nussinov R. 2003. Extended disordered proteins: targeting function with less scaffold. *Trends Biochem Sci* **28**: 81–85.
- Gunster MJ, Satijn DP, Hamer KM, den Blaauwen JL, de Bruijn D, Alkema MJ, van Lohuizen M, van Driel R, Otte AP. 1997. Identification and characterization of interactions between the vertebrate polycomb-group protein BMI1 and human homologs of polyhomeotic. *Mol Cell Biol* **17**: 2326–2335.
- Hansen JC, Lu X, Ross ED, Woody RW. 2006. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J Biol Chem* **281**: 1853–1856.
- Hashimoto N, Brock HW, Nomura M, Kyba M, Hodgson J, Fujita Y, Takihara Y, Shimada K, Higashinakagawa T. 1998. RAE28, BMI1, and M33 are members of heterogeneous multimeric mammalian Polycomb group complexes. *Biochem Biophys Res Commun* **245**: 356–365.
- Hemenway CS, Halligan BW, Levy LS. 1998. The Bmi-1 oncoprotein interacts with dinG and MPh2: the role of RING finger domains. *Oncogene* **16**: 2541–2547.
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, et al. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**: D211–D215. doi: 10.1093/nar/gkn785.
- Ingham PW, Whittle JRS. 1980. Trithorax: a new homeotic mutation of *Drosophila melanogaster* causing transformations of abdominal and thoracic imaginal segments. *Mol Gen Genet* **179**: 607–614.
- Kennison JA. 1993. Transcriptional activation of *Drosophila* homeotic genes from distant regulatory elements. *Trends Genet* **9**: 75–79.
- Kennison JA. 1995. The Polycomb and trithorax group proteins of *Drosophila*: trans-regulators of homeotic gene function. *Annu Rev Genet* **29**: 289–303.
- Kennison JA, Tamkun JW. 1988. Dosage-dependent modifiers of polycomb and antennapedia mutations in *Drosophila*. *Proc Natl Acad Sci* **85**: 8136–8140.
- King IF, Francis NJ, Kingston RE. 2002. Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro. *Mol Cell Biol* **22**: 7919–7928.
- King IF, Emmons RB, Francis NJ, Wild B, Muller J, Kingston RE, Wu CT. 2005. Analysis of a polycomb group protein defines regions that link repressive activity on nucleosomal templates to in vivo function. *Mol Cell Biol* **25**: 6578–6591.
- Kuroski MA, Bujnicki JM. 2003. GeneSilico protein structure prediction meta-server. *Nucleic Acids Res* **31**: 3305–3307.
- Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**: 2893–2905.
- Kwon H, Imbalzano AN, Khavari PA, Kingston RE, Green MR. 1994. Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* **370**: 477–481.
- Lavigne M, Francis NJ, King IF, Kingston RE. 2004. Propagation of silencing; recruitment and repression of naive chromatin in trans by polycomb repressed chromatin. *Mol Cell* **13**: 415–425.
- Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, et al. 2006. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**: 301–313.
- Lewis EB. 1947. New mutants: report of P. Lewis. *Drosoph Inf Serv* **21**: 69.
- Lewis EB. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**: 565–570.
- Lo SM, Francis NJ. 2010. Inhibition of chromatin remodeling by polycomb group protein posterior sex combs is mechanistically distinct from nucleosome binding. *Biochemistry* **49**: 9438–9448.
- Lo SM, Ahuja NK, Francis NJ. 2009. Polycomb group protein Suppressor 2 of zeste is a functional homolog of Posterior Sex Combs. *Mol Cell Biol* **29**: 515–525.
- Ma J, Ptashne M. 1987. A new class of yeast transcriptional activators. *Cell* **51**: 113–119.
- Min J, Zhang Y, Xu RM. 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* **17**: 1823–1828.
- Morey L, Helin K. 2010. Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci* **35**: 323–332.
- Muller J. 1995. Transcriptional silencing by the Polycomb protein in *Drosophila* embryos. *EMBO J* **14**: 1209–1220.
- Muller J, Verrijzer P. 2009. Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr Opin Genet Dev* **19**: 150–158.
- Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. 2002. Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**: 197–208.
- Neira JL, Roman-Trufero M, Contreras LM, Prieto J, Singh G, Barrera FN, Renart ML, Vidal M. 2009. The transcriptional repressor RYBP is a natively unfolded protein which folds upon binding to DNA. *Biochemistry* **48**: 1348–1360.

- Pearce JJ, Singh PB, Gaunt SJ. 1992. The mouse has a Polycomb-like chromobox gene. *Development* **114**: 921–929.
- Ren X, Vincenz C, Kerppola TK. 2008. Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. *Mol Cell Biol* **28**: 2884–2895.
- Sandhu KS. 2009. Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins. *J Mol Recognit* **22**: 1–8.
- Satijn DP, Otte AP. 1999. RING1 interacts with multiple Polycomb-group proteins and displays tumorigenic activity. *Mol Cell Biol* **19**: 57–68.
- Schoorlemmer J, Marcos-Gutierrez C, Were F, Martinez R, Garcia E, Satijn DP, Otte AP, Vidal M. 1997. Ring1A is a transcriptional repressor that interacts with the Polycomb-M33 protein and is expressed at rhombomere boundaries in the mouse hindbrain. *EMBO J* **16**: 5930–5942.
- Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. 2007. Genome regulation by polycomb and trithorax proteins. *Cell* **128**: 735–745.
- Schwartz YB, Pirrotta V. 2008. Polycomb complexes and epigenetic states. *Curr Opin Cell Biol* **20**: 266–273.
- Schwartz YB, Kahn TG, Nix DA, Li XY, Bourgon R, Biggin M, Pirrotta V. 2006. Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat Genet* **38**: 700–705.
- Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE. 1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**: 37–46.
- Shoemaker BA, Portman JJ, Wolynes PG. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci* **97**: 8868–8873.
- Sif S, Stukenberg PT, Kirschner MW, Kingston RE. 1998. Mitotic inactivation of a human SWI/SNF chromatin remodeling complex. *Genes Dev* **12**: 2842–2851.
- Sif S, Saurin AJ, Imbalzano AN, Kingston RE. 2001. Purification and characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes. *Genes Dev* **15**: 603–618.
- Sigler PB. 1988. Transcriptional activation. Acid blobs and negative noodles. *Nature* **333**: 210–212.
- Simon JA, Kingston RE. 2009. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**: 697–708.
- Slifer EH. 1942. A mutant stock of *Drosophila* with extra sex-combs. *J Exp Zool* **90**: 31–40.
- Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**: 207–234.
- The UniProt Consortium. 2011. Ongoing and future developments at the universal protein resource. *Nucleic Acids Res* **39**: D214–D219. doi: 10.1093/nar/gkq1020.
- Utlely RT, Ikeda K, Grant PA, Cote J, Steger DJ, Eberharder A, John S, Workman JL. 1998. Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature* **394**: 498–502.
- Uversky VN, Dunker AK. 2010. Understanding protein non-folding. *Biochim Biophys Acta* **1804**: 1231–1264.
- Wang R, Ilangovan U, Robinson AK, Schirf V, Schwarz PM, Lafer EM, Demeler B, Hinck AP, Kim CA. 2008. Structural transitions of the RING1B C-terminal region upon binding the polycomb cbox domain. *Biochemistry* **47**: 8007–8015.

Appendix B: Functional dissection of Polycomb repressive complex 1 reveals the importance of a charged domain

This paper was published in the Cold Spring Harbor Symposia on Quantitative Biology online April 18, 2010.

Grau DJ, Antao JM, Kingston RE. 2010. Functional dissection of Polycomb repressive complex 1 reveals the importance of a charged domain. *Cold Spring Harb Symp Quant Biol* 2010 **75**: 61-70.

Functional Dissection of Polycomb Repressive Complex 1 Reveals the Importance of a Charged Domain

D.J. GRAU,^{1,2,4} J.M. ANTÃO,¹⁻⁴ AND R.E. KINGSTON^{1,2}

¹Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114;

²Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115; ³Gulbenkian PhD Program in Biomedicine, Instituto Gulbenkian de Ciência (IGC), Oeiras, Portugal

Correspondence: Kingston@molbio.mgh.harvard.edu

Silencing of homeotic genes requires the Polycomb repressive complex 1 (PRC1) family of protein complexes, which are composed of Polycomb-group (PcG) proteins and frequently include other subunits. We discuss here two aspects of PRC1 that might contribute to this activity. Inhibiting the action of remodeling factors via chromatin compaction is believed to be one mechanism by which PRC1 represses genes. We show that PRC1s from fly and mouse have conserved this activity as complexes. Additionally, we provide evidence that a different subunit in the mouse complex retains the conserved repression activity and that activity appears to be mediated by charge interactions. We show that Zeste interacts specifically with the Ph subunit of PRC1 and discuss the possibility of these factors contributing to spreading of PRC1 complexes. Our results suggest that one aspect of PRC1 repression is likely to be mediated by charge–charge interactions.

A key aspect of development is the ability to maintain master regulatory genes in a repressed state when appropriate. Misexpression of even a single master regulatory gene, such as those encoded in the HOX loci, can cause a cell to behave in a manner incompatible with its body location and tissue type. The most prominent set of factors responsible for the maintenance of a repressed state at master regulatory genes is called the Polycomb group (PcG), after the founding gene in this family, Polycomb, discovered in *Drosophila* in the late 1940s (Lewis 1947; Lewis 1978). The PcG genes, which number roughly 16 depending on species, form several protein complexes that are involved in repression. One of these complexes, Polycomb repressive complex 1 (PRC1), is believed to be the central engine of repression.

The mechanisms via which the PcG system maintains a repressed state faithfully throughout the lifetime of a cell lineage are not fully understood. The system is targeted to genes by specific loci called Polycomb response elements (PREs). Several distinct DNA-binding factors bind to PREs, localize PcG complexes via interactions with these complexes, and allow these complexes to act at adjacent, and many times distant (>50 kb), regions of the genome (for review, see Ringrose and Paro 2007). More recently, there have been some suggestions that noncoding RNAs (ncRNAs) are also involved in targeting (Rinn et al. 2007; Zhao et al. 2008). Once the targets are found, repression involves methylation of lysine 27 of histone H3, ubiquitylation of lysine 119 of histone H2A, binding to and stabilization of nucleosome structure, and direct inhibition of the transcription machinery (for review, see Simon and

Kingston 2009). Given the number of PcG components, the variety of genes that are regulated, and the fact that key targets such as the HOX clusters occupy ~500 kb in the mammalian and in the *Drosophila* genomes, it seems likely that a variety of mechanisms act in distinct combinations on the various targets.

Several complexes are formed by PcG gene products. In *Drosophila*, these include PRC1 (Shao et al. 1999; Saurin et al. 2001), PRC2 (Czermin et al. 2002; Kuzmichev et al. 2002; Müller et al. 2002), PhoRC (Klymenko et al. 2006), and PR-DUB (Scheuermann et al. 2010). PhoRC binds to many PREs to help target the repressive PcG machinery; PRC2 methylates lysine 27 of histone H3 to help target binding by the PRC1 complex, which creates a repressive state on chromatin; and PR-DUB deubiquitylates histone H2A, thus counteracting one of the functions of the PRC1 family of complexes. The extent to which this balance of counteracting PcG activities drives a repressive state is not understood currently. It is also not clear whether PRC2 is involved only in targeting PRC1 or whether it directly represses gene expression in coordination with PRC1 as well.

PRC1 family members have been found in complexes distinct from PRC1 as originally defined in *Drosophila* (Table 1). These complexes also have tight interactions with proteins that do not display a classic PcG phenotype and thus are not bona fide PcG members. Complexes in this family have several functions. Complexes that contain the Bmi-1 and Ring1B proteins are able to ubiquitylate histone H2A; the most active of these complexes also contain the KDM2B protein or its homolog (Gearhart et al. 2006; Lagarou et al. 2008). The role for ubiquitylation in directing repression is under investigation; it has been proposed to impede transcriptional elongation (Stock et al. 2007) but does not appear key for repression of HOX loci in mice

⁴These authors contributed equally to this work.

Table 1. Selected PcG proteins

<i>Drosophila</i>	Mouse	Complex	Domains	Functional role(s)
Enhancer of zeste (E(sz))	Ezh1P Ezh2 (KMT6, Enx-1)	PRC2 PRC2	SET SET	H3K27 Mouse
Suppressor zeste 12 (Su(z)12)	Suz12	PRC2		Enhances Mtase activity of E(z)
Extra sex combs (Esc)	Eed (lusk)	PRC2	WD40 repeats	Enhances Mtase activity of E(z)
Nucleosome remodeling factor 55 (Nurf55/Caf-1)	Rbbp4 (Caf-1, mRbAp48)	PRC2	WD40 repeats	Nucleosome binding
Polycomb-like (Pcl)	Phf1 (Pcd1) Mtf2 (Pcl2) Phf19 (Pcl3)	Interacts with PRC2	PHD fingers and Tudor domain	Recruitment, enhances Mtase activity of PRC2
Polycomb (Pc)	Cbx2 (M33, MOD2, Pc1) Cbx4 (Pc2) Cbx6 Cbx7 Cbx8 (Pc3)	PRC1 PRC1 PRC1 PRC1 PRC1	Chromodomain AT-hook and Pc-box Chromodomain and Pc-box Chromodomain and Pc-box Chromodomain and Pc-box Chromodomain and Pc-box	Binds H3K27triMe Binds H3K27triMe Binds H3K27triMe Binds H3K27triMe Binds H3K27triMe
Posterior sex combs (Psc) and Suppressor of zeste 2 (Su(z)2)	Bmi1 (Pcgf4) Mel18 (Pcgf2, Rnf110) Pcgf1 (Nspc1)	PRC1 PRC1 BCOR	RING finger RING finger RING finger	Enhances ubiquitination Compacts nucleosomes
Polyhomeotic proximal (Ph-p) and distal (Ph-d)	Phc1 (Edr, Mph1, Rae-28) Phc2 (Edr2, Mphj2, p36) Phc3 (Edr3, Hph3)	PRC1 PRC1 PRC1	SAM SAM SAM	Spreading? Spreading? Spreading?
Sex combs extra (Sce/dRing)	Ring1 (Ring1A) Rnf2 (Ring1B, dinG)	PRC1/BCOR PRC1/BCOR	RING finger RING finger	E3 ligase E3 ligase
Pleiohomeotic (Pho)	YY1 (NF-E1)	PhoRC	Zn fingers	Recruitment?
dSfmbt	mSfmbt (Smr)	PhoRC	Zn finger, MBT, and SAM	Binds methylated histones
Calypso	Bap1	PR-DUB	UCH	Deubiquitinates H2A
Additional sex combs (Asx)	Asx11 Bcor	PR-DUB BCOR	Ankyrin repeat	Enhances Calypso activity Targeting H2A ubiquitination?
dKDM2	Kdm2b (Fbx110)	BCOR	CXXC, PHD finger, F-box, and leucine-rich repeats	H3K36 demethylase
dRYBP	Rybp (DEDAF, YEAF1)	BCOR	Zn finger	Protein-protein interaction
skpA	Skp1a (EMC19, OCP2, SKP1A, p19A)	BCOR	SKP1	

(Eskeland et al. 2010). It is possible that the importance of ubiquitylation varies according to the repressed target gene. A second activity of the PRC1 family complexes, and a focus of this chapter, is their ability to create a compacted chromatin state that is refractory to ATP-dependent remodeling. Creation of a compacted state has been proposed to direct repression by blocking steps in the transcription process (Nakagawa et al. 2008), possibly at the level of transcription initiation or transcription elongation. A second focus of this chapter is the ability of auxiliary factors to modulate activity and/or targeting of PRC1 activities.

To dissect the mechanism of PRC1 function, we have analyzed domains of some of its central components. In previous work, we showed that the Psc protein is responsible for creating a compacted state of the template (Francis et al. 2001) and defined the regions of Psc responsible

for that activity (King et al. 2005). An issue that arose was that there was no obvious homologous domain in mammalian Psc homologs: If this activity is central to PRC1 function, where might a domain reside that performs this activity in mammals? In previous work, we had also shown that a key protein involved in both activation and repression in *Drosophila*, the Zeste protein, interacts directly with the core components of PRC1 (Saurin et al. 2001; Mulholland et al. 2003). This was perplexing because, although this protein is known to have roles in regulation in *Drosophila*, it has been shown to display phenotypes consistent with the Trithorax group (Judd 1995), a set of genes isolated by their ability to suppress PcG phenotypes. We have addressed the ability of Zeste to dock with PRC1 and tested whether there are functional outcomes from this interaction.

MATERIALS AND METHODS

Protein Expression

Protein expression was performed as described previously (Phelan et al. 1999). Briefly, *Sf9* cells in the exponential growth phase were infected with baculovirus for the desired proteins, and cells were harvested 40 h after infection. Nuclear extracts were prepared as described (Abmayr et al. 2001), and protein was bound to an M2 affinity resin (Sigma), washed with BC buffer (20 mM HEPES at pH 8.0, 0.2 mM EDTA, 20% glycerol, 0.2 mM PMSF, 0.5 mM DTT), with KCl concentrations up to 2 M, and eluted with 0.4 mg/mL Flag peptide.

Protein-Protein Interaction Assays

Sf9 cells at a density of 5×10^5 /mL were coinfecting with baculovirus for the candidate interacting proteins. Cells were harvested after 40 h and washed with PBS, and cell extracts were prepared by freezing/thawing three times in the presence of protease inhibitors. Extracts were incubated with M2 beads, washed with BC buffer with KCl concentrations up to 2 M, and eluted with 0.4 mg/mL Flag peptide. Proteins were separated by SDS-PAGE and detected either by Western blot or Colloidal Blue staining (Invitrogen).

Generation of Nucleosomal Arrays

The ClaI/Asp118 fragment of pG5E4 containing 10 5S nucleosome positioning sequences was purified using standard molecular biology techniques. Purified G5E4 fragment was end-labeled using Klenow and [α - 32 P]dATP. Core nucleosomes were prepared from HeLa nuclei and assembled into nucleosomal arrays using the salt gradient dialysis method as previously described (Sif et al. 2001).

Restriction Enzyme Accessibility Assay

The restriction enzyme accessibility assay was performed essentially as described (Francis et al. 2001). Briefly, 1.5 nM nucleosomes was incubated with P_oG proteins at the indicated concentration for 30 min at 30°C. Afterward, hSWI/SNF was added at concentrations determined not to be rate-limiting, in the presence of 8 units of HhaI (New England BioLabs) in 20 μ L of reaction buffer containing 12 mM HEPES (pH 7.9), 0.2 mM EDTA, 1.5 mM MgCl₂, 2.4 μ g of BSA, 5% glycerol, 2 mM ATP, and 1 mM DTT. After an additional 1 h at 30°C, the 20- μ L reactions were quenched by the addition of 10 μ L of stop buffer: 10 mM Tris (pH 7.7), 35 mM EDTA, 1% SDS, 0.1 mg/mL bromophenol blue, and 1.5 mg/mL Proteinase K. Stopped reactions were incubated for 45 min at 55°C, then resolved on a 1% agarose gel before exposure to a phospho-imaging screen and quantification using a Typhoon PhosphorImager and ImageQuant software (GE Healthcare).

Evolutionary Analysis

The UniProt database was queried for proteins containing either chromodomains or RING domains. These lists of proteins were further filtered by removing proteins that

did not contain a C-box in the chromodomain list or did not contain the extra region of homology in the RING protein set. The proteins in these lists were classified as active or inactive based on predicted overall charge, with +10.2 being the cutoff point, with proteins more positively charged being classified as active. The cladogram was generated by aligning the 18S rRNA sequences for each of the listed organisms, and the cladogram was drawn using the neighbor-joining method using NJplot.

RESULTS

Chromatin Compaction Is Conserved in the Mouse PRC1 Complex

The domains responsible for chromatin compaction by *Drosophila* PRC1 were shown by solution studies and by electron microscopy to reside in the carboxy-terminal two-thirds of the Psc protein (Francis et al. 2004; King et al. 2005). Function is dispersed in a large region of the protein, and there is a good correlation between how mutations in Psc affect the in vivo phenotype and how those mutations affect compaction by Psc in vitro. Bmi1, one of the mouse homologs of Psc, is structurally quite different from Psc. Although Bmi1 and Psc share a conserved RING finger and HTH domain, Bmi1 lacks the extended carboxy-terminal region of Psc that is required for both in vitro and in vivo function. Thus, we were interested in whether Bmi1 or another mouse PRC1 protein was capable of functioning in isolation in vitro in a manner similar to Psc.

We expressed and purified individual components of mouse Polycomb core complex (mPCC) and tested them for the ability to inhibit remodeling in a solution protocol based on restriction enzyme accessibility. Restriction enzyme cleavage is inhibited when the target DNA is organized into a nucleosome. ATP-dependent remodeling proteins modify the nucleosomal structure in ways that allow accessibility and cleavage, but in the presence of PRC1, this activity is blocked. We measure the ability of a PRC1 preparation to function by titrating it into a remodeling reaction and measuring the extent to which restriction enzyme access is inhibited. This function has generally correlated well with the ability to compact chromatin, as studied using electron microscopy (Francis et al. 2004). Using this protocol, a reconstituted complex containing proteins Bmi1, M33/Cbx2, and Ring1a showed activity similar to that seen with *Drosophila* Psc (Fig. 1).

We tested individual mouse P_oG proteins to determine whether any single protein is able to inhibit remodeling. Bmi1 does not significantly inhibit remodeling at the concentrations of protein tested (Fig. 1), consistent with the fact that it does not contain a region homologous to the compaction domain defined in Psc. Next, we tested the other components of the minimal core complex for repression activity. Similar to Bmi1, Ring1a does not show any appreciable activity on its own. However, M33/Cbx2 inhibits remodeling to an extent similar to the mouse core complex and Psc. This is surprising because we had previously observed little in vitro activity of the *Drosophila* homolog of M33/Cbx2, Pc.

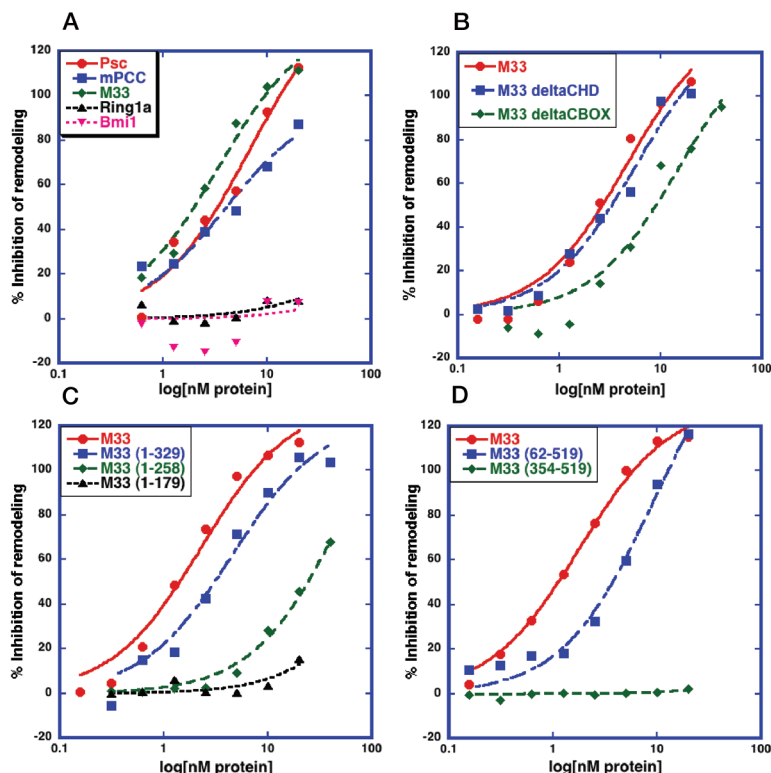


Figure 1. Repressive activity of PcG protein M33. PcG proteins were titrated into a reaction containing nucleosomal arrays and incubated before addition of hSWI/SNF and HhaI. Remodeling is measured by quantifying the ratio of DNA cut by HhaI to uncut DNA. The percent inhibition of remodeling is calculated by using the following equation:

$$\frac{(\% \text{ uncut with hSWI/SNF and PRC1 or PCC} - \% \text{ uncut with hSWI/SNF})}{(\% \text{ uncut without hSWI/SNF} - \% \text{ uncut with hSWI/SNF})} \times 100.$$

(A) The ability of M33 to inhibit remodeling is similar to Psc and mPCC. (B) Activity of M33 domain deletions. (C) Activity of M33 carboxy-terminal truncations. (D) Activity of M33 amino-terminal truncations.

To further investigate the observed activity of M33/Cbx2, we performed a structure/function analysis to determine what domain(s) of M33/Cbx2 are required for functional activity. Deletion of either the chromodomain or the C-box of M33/Cbx2 does not significantly decrease its activity (Fig. 1B,C). The chromodomain of Cbx/Pc proteins is involved in targeting to chromatin (Fischle et al. 2003), and the C-box is required for repression activity in transfection assays (Schoorlemmer et al. 1997) as well as forming interactions with other PRC1 components (Satijn et al. 1997). Targeting is not anticipated to be necessary in vitro because association with the template is driven by mass action; thus, the lack of requirement for the chromodomain was not surprising. The lack of requirement for the C-box in vitro indicates that this domain might function primarily by directing interactions with other PcG proteins and thus may not be necessary for function of the isolated protein.

We next decided to create a series of truncation mutants to map where the repression activity was located. Progressive deletion into the core of M33/Cbx2 from either end of the protein results in increasing loss of in vitro activity (Fig. 1C,D). This observation is reminiscent of what was seen for Psc: Functional activity spreads throughout the extended carboxyl terminus of Psc, with further truncation

of the carboxyl terminus correlating with loss of both in vitro and in vivo activity. This is consistent with the idea that there is no well-defined repression domain but, rather, that repression activity is spread throughout the carboxyl terminus of the protein.

Through examination of the M33/Cbx2 sequence, we observed what appears to be an overrepresentation of two basic amino acids, lysine and arginine. This abundance of basic amino acids is similar to what is found in the carboxyl terminus of Psc. We hypothesized that one of the repressive activities of the PcG proteins is mediated by a region of a PRC1 subunit with a high overall positive charge. We imagined that a localized region of high positive charge could potentially interact with nucleosomal DNA and mediate the compaction of chromatin, and block transcriptional activators from having access to their cognate sequences.

If this hypothesis were true, we would expect there to be a correlation between the charge of a given M33/Cbx2 mutant and resulting in vitro activity. In fact, if we plot the charge of a given M33/Cbx2 mutant versus the concentration required for 50% inhibition of remodeling (IC_{50}), we see a significant correlation (Fig. 2). This result is consistent with the idea that charge is an important determinant of PRC1 activity.

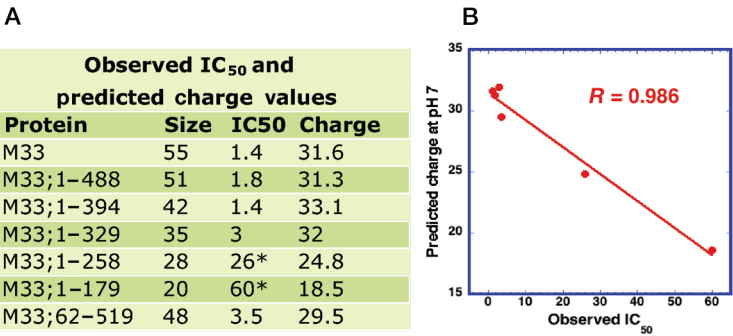


Figure 2. Correlation of M33 repression activity and charge. (A) Table lists charge and observed activity of M33/Cbx2 mutants. “Size” column is in kilodaltons. Charge was estimated at pH 7.0 using lasergene software. (Asterisks) IC₅₀ was inferred by extrapolating curve fits. (B) Plot of predicted charge at pH 7 and observed IC₅₀ for inhibition of remodeling.

Evolutionary Divergence of PRC1 Function

Interestingly, the above analysis suggests that a different subunit of mouse core PRC1 has coopted the activity of Psc from the *Drosophila* core complex. Bmi1, the mouse homolog of Psc, is inactive in our in vitro assay. This is in contrast to M33/Cbx2, a mouse homolog of Pc, which is active. These observations could be explained in an evolutionary context by either coevolution of this particular PcG activity or by gain of this activity by one subunit followed by loss in another. Regardless of what actually occurred evolutionarily, both of these models suggest that organisms evolutionarily close to *Drosophila* are expected to have Psc homologs with high overall positive charge. Likewise, we expect organisms evolutionarily closer to mouse to have Pc homologs with high overall positive charge. To test this hypothesis, we examined divergent homologs of Psc and Pc in multiple species and analyzed the charge characteristics of each protein.

To accomplish this in an unbiased manner, we used a computational approach to “call” PcG proteins and classify them as active or inactive based on charge. To call Pc (or M33) homologs, we queried the SWISS-PROT database for proteins containing a chromodomain and a C-box motif, the hallmarks of this class of PcG protein. To find Psc homologs, the database was queried using the conserved RING domain plus the extended homology domain of Psc. This method yielded 44 M33/Cbx2 homologs and 59 Psc homologs. Although we cannot exclude the possibility that putative PcG proteins were missed, the proteins

that were called should be representative and are likely to contain the majority of homologs in the database. Notably, many proteins that were called have GO classifications of PcG proteins, and one protein, *Caenorhabditis elegans* Mig-32, has recently been shown to be a functional Psc homolog in the worm (Karakuzu et al. 2009).

To classify the proteins as either active or inactive, we used the predicted charge at pH 7 and set a threshold of +10.2 for candidate homologs to be categorized as inactive. This cutoff was defined based on results with tested PcG proteins as described above. Even though we expect the activity to fall along a range, dependent on the charge level, this binary system makes testing the predictions simpler. Using this criterion, we find that 32 of 44 M33/Cbx2 homologs are predicted to be “active” in our assay. In contrast, 55 of 59 Psc homologs are predicted to be “inactive.”

To further understand these results in the context of evolution, we generated a cladogram using an alignment of the 18S rRNAs from each of the species in which we found putative PcG proteins (Fig. 3). There is a clear dichotomy: Organisms more closely related to *Drosophila* have Psc homologs that are predicted to be active, whereas the organisms more closely related to mouse have Pc (M33/Cbx2) homologs that are predicted to be active. Of note is that despite the high number of Psc homologs in mammals, none was classified as active. These results support the hypothesis that during evolution this particular activity of PRC1 was “swapped” among different subunits.

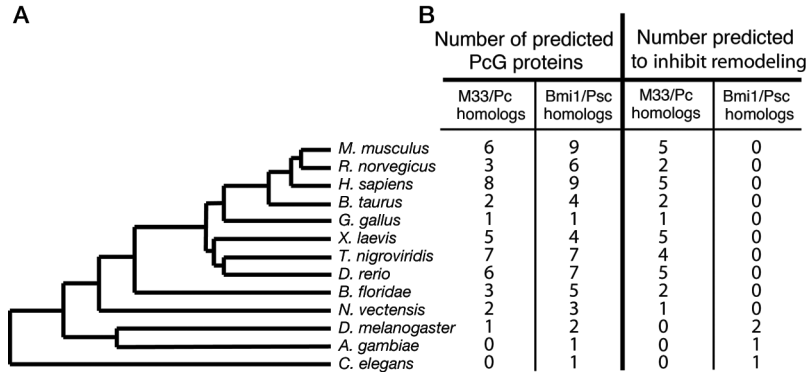


Figure 3. Predicted repression activity of evolutionary PcG homologs. (A) Cladogram of species generated by alignment of 18S ribosomal RNA sequences. (B) Number of predicted PcG proteins from each species and number of proteins over the charge threshold of +10.2.

Interaction of the Zeste Protein with PRC1 in *Drosophila*

The targeting of Polycomb-group proteins and the activity at loci distant from the nucleation sites are fundamental issues in the mechanism of PcG silencing. Simple models of sequential action by PRC2, which methylates K27 on histone H3, and PRC1, which binds the K27 mark via the Pc chromodomain, are not sufficient to explain PRC1 recruitment. The affinity of the chromodomain of Pc for trimethylated lysine 27 on histone H3 is low (Fischle et al. 2003), and certain Pc homologs do not display a strong preference for K27 methylation over K9 methylation (Bernstein et al. 2006). Histone marks might confer an extra level of stability to the binding of PRC1, which is important, but histone methylation is unlikely to be the sole mechanism of targeting.

Studies using *Drosophila* have identified a multitude of DNA-binding proteins that recognize PREs and are proposed to recruit PcG complexes via direct interaction. One such factor, Zeste (Fig. 4A), was identified as an integral

component of the PRC1 complex as isolated from *Drosophila* embryos (Saurin et al. 2001), which suggested a mechanism for directly targeting the complex to chromatin. Interestingly, Zeste, which is a transcriptional activator for the *Ubx* gene (Biggin et al. 1988), has been classified as a Trithorax-group protein and shown to interact physically with several Brahma-associated factors (BAFs) in the Brahma complex (Kal et al. 2000). Such dualism is confirmed by the fact that Zeste is required for the maintenance of both active (Déjardin and Cavalli 2004) and repressed (Hur et al. 2002) states of transcription, in both cases, an activity that depends on TrxG/PcG function.

To characterize how the Zeste protein interacts with PRC1, we used a reconstitution approach to define the binding interaction and to determine whether we could measure a functional outcome of that interaction. We coinfecting *Sf9* cells with baculovirus for the four core PRC1 factors, or these factors plus Zeste, and purified the PCC and PCCZ complexes, respectively, through the Flag tag on the subunit Ph. We were able to generate complexes with stoichiometric levels of Zeste protein, indicating that

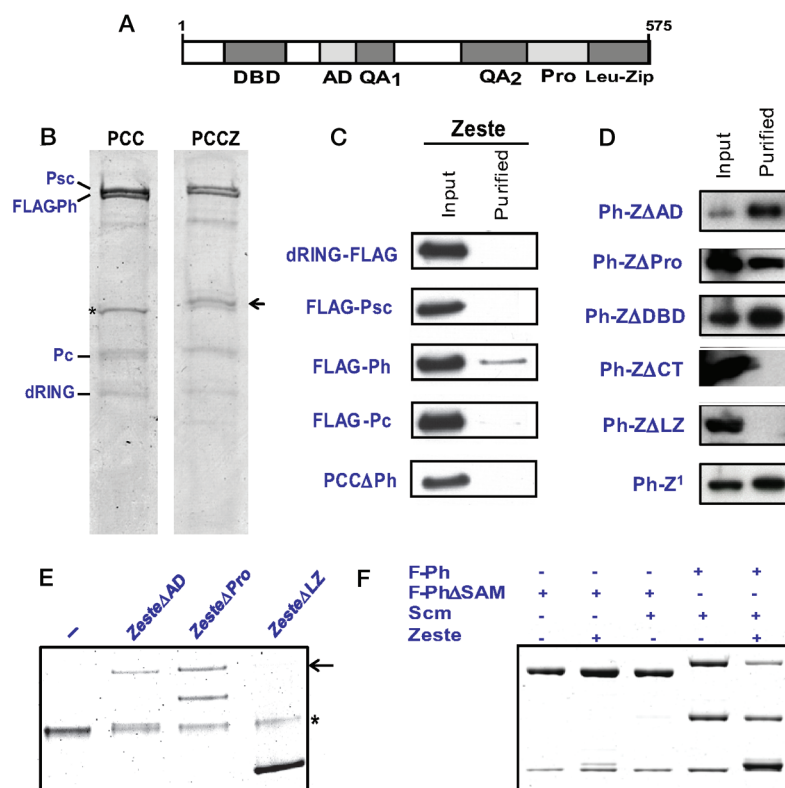


Figure 4. The interaction of Zeste with Polycomb. (A) The domain structure of the Zeste protein. (DBD) DNA-binding domain, (AD) activation domain, (QA) glutamine/alanine-rich domain, (Pro) proline-rich domain, (Leu-Zip) leucine zipper. (B) Purification of recombinant PCC and PCC-Zeste from Baculovirus-infected *Sf9* cells. (Arrow) Zeste, (asterisk) contaminating band. (C) Western blot for Zeste, for input and protein purified through the Flag tag on the indicated protein, coinfecting with Zeste. PCCΔPh is a coinfection with FLAG-Psc, dRING, Pc, and Zeste. (D) FLAG-PH was coinfecting with the indicated mutants of Zeste and purified through the Flag tag. Western blot for input and purified protein for Zeste, (Z¹) K425M point mutant, (ZΔCT) carboxy-terminal truncation that eliminates the proline-rich and leucine-zipper domains. (E) Full-length HA-tagged Zeste was coinfecting with a Flag-tagged version of the indicated truncation mutants, and protein that was purified through the Flag tag was detected by Colloidal Blue staining. (Arrow) Full-length Zeste, (asterisk) nonspecific protein, (– lane [above, left]) infection with full-length Zeste alone. (F) Colloidal Blue staining of protein purified through the Flag tag of PH, in the combinations indicated. Zeste is the thicker band just above the nonspecific protein found on all purifications.

Zeste associates strongly with the core PRC1 (Fig. 4B). We then asked whether this interaction is mediated by a direct physical contact with one of the core factors. To this end, we coinfecting *Sf9* cells with baculovirus for Zeste and a Flag-tagged version of each of the PCC proteins. We then purified the Flag-tagged protein and identified Zeste in the eluates by western blot. Ph is both sufficient and necessary for the association of Zeste with PCC, because Ph alone bound to Zeste and a partial PCC complex lacking Ph (PCCΔPh) was unable to interact with Zeste (Fig. 4C).

Having defined Ph as the interacting protein, we used the same strategy described above to characterize the domains required for interaction between Ph and Zeste. We made several deletion mutants of Zeste for use in interaction studies. Mutants lacking the activation domain, proline-rich domain, DNA-binding domain, or a K425M mutant (the *z'* mutant, which displays a neomorphic phenotype in flies) were all able to interact with FLAG-Ph (Fig. 4D). In contrast, mutants that lack the carboxyl terminus (proline-rich and leucine-zipper domains) or that lack the leucine zipper only, do not copurify with FLAG-Ph.

The leucine-zipper domain of Zeste has previously been found to be involved in a physical interaction with Moira, a subunit of the TrxG Brahma complex (Kal et al. 2000). Additionally, it has been implicated in aggregation of the Zeste protein, which has been suggested to have a role in the in vivo function of Zeste (Chen and Pirrotta 1993). Consistent with this observation, a mutant lacking the leucine zipper was shown to be deficient in self-association (Fig. 4E). We tested self-association by coinfecting *Sf9* cells with full-length HA-tagged Zeste and a Flag-tagged version of a Zeste mutant that lacked one of either the activation, the proline-rich, or the leucine-zipper domains. Protein purified over an M2 Flag affinity column was separated by PAGE and stained with Colloidal Blue. The leucine-zipper mutant is unable to recruit full-length Zeste, indicating that this domain is necessary for self-interaction.

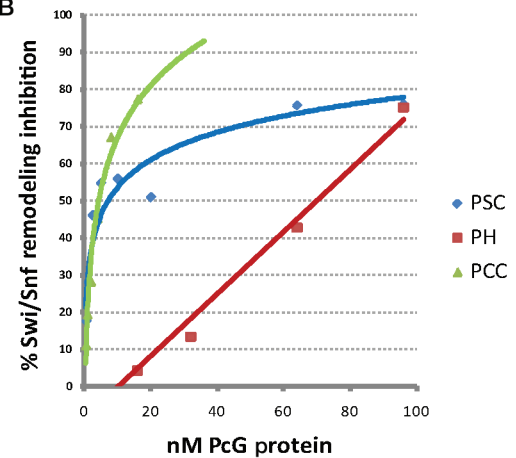
We next were interested in testing whether a known interaction module in Ph, the SAM domain, mediates the interaction with Zeste. We coinfecting *Sf9* cells with FLAG-Ph or FLAG-PhΔSAM and Zeste. As a positive control, we used Sex Comb on the Midleg (Scm), a PcG protein known to interact with the SAM domain of Ph. As expected, we could see a SAM-dependent interaction between Ph and Scm (Fig. 4F). Furthermore, the PhΔSAM mutant had a much lower affinity for Zeste, suggesting that the SAM domain is a common docking site for both Scm and Zeste. Binding to Ph is possibly done using different surfaces of the SAM domain, because both Scm and Zeste can copurify with Ph simultaneously.

We wished to determine whether the interaction between Ph and Zeste has a measurable impact on activity of either Ph alone or the core PRC1 complex. To perform these studies, we first determined the function of Ph alone. As discussed above, the inhibitory activity of PRC1 on chromatin appears to be due to the high positive charge of the complex, a characteristic that is unique to PRC1 among the known fly PcG complexes (Fig. 5A). Interestingly, a different complex that includes Psc, the dRAF complex (Lagarou et al. 2008), is not known to have a di-

A

Drosophila PRC1		Mouse PRC1	
Protein	Charge	Protein	Charge
Posterior sex combs	82.1	BMI1	2.2
Polyhomeotic	25.8	RAE-28	18.3
Polycomb	4.2	M33	32.5
dRING	0.6	RING1B	-1.1
dPCC	112.7	mPCC	51.9
Zeste	-2.7		
PRC2	-17		
PhoRC	-31.2		
dRAF	-17.3		

B



C

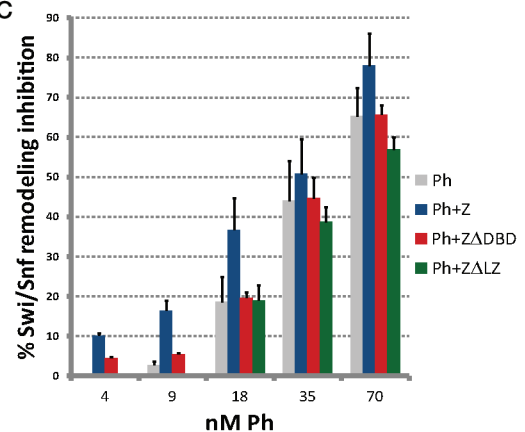


Figure 5. Functional analysis of Polyhomeotic and Zeste. (A) Charge of the PRC1 core complex proteins from *Drosophila* and mouse and of other *Drosophila* PcG complexes and Zeste. (B) Swi/Snf remodeling inhibition by dPCC, Psc, and Ph. (C) Interaction of wild-type and mutant Zeste with Ph in Swi/Snf.

rect impact on chromatin remodeling activity and bears an overall charge of -17.3 . This suggests that other factors in PRC1 contribute to the overall charge and, therefore, activity. Ph, although not as positively charged as Psc, bears a net charge above the threshold that was used for calling a PcG protein as potentially active in the inhibition of remodeling. Likewise, a Ph homolog in mouse, RAE-28, also has an intermediate positive charge.

To determine whether the predicted charge of these proteins was indicative of biochemical activity, we measured

the ability of purified recombinant PCC, Psc, and Ph to inhibit remodeling with the REA assay. As expected, PCC has the strongest inhibitory activity, closely followed by Psc (Fig. 5B). Ph displayed measurable inhibitory activity, albeit only at higher concentrations than PCC or Psc.

Zeste has previously been shown to have an impact on the activity of PCC (Mulholland et al. 2003). Given the physical interaction between Zeste and Ph (Fig. 4), Zeste might directly enhance the activity of Ph. To test this hypothesis, we again used the REA assay. By preincubation of the nucleosomal arrays with either Ph alone or Ph plus wild-type or mutant versions of Zeste, we could determine whether Zeste enhances the activity of Ph (Fig. 5C). The inclusion of Zeste in the preincubation has a measurable, but modest, effect on the activity of Ph, more prominent at lower concentrations. Removal of either the DNA-binding domain (ZADBD) or the leucine-zipper domain (ZALZ) abrogated the stimulatory effect of Zeste. These results suggest that the interaction of Zeste with Ph stimulates the inhibitory activity of Ph, dependent on both the DNA-binding domain and the leucine-zipper domain of Zeste. This dependence might be due to either the reduced physical interaction of the ZALZ mutant with Ph or to the inability of this mutant to oligomerize. The effect of Zeste on Ph is unlikely to be related to charge because Zeste has a net negative charge.

DISCUSSION

We have examined the biochemically active regions of two key components of PRC1 and have generated the hypothesis that the high net positive charge of these regions is important for the ability of PRC1 to compact chromatin. Surprisingly, the region enriched for basic amino acids, which was originally identified in the *Drosophila* Psc protein, appears to have swapped location to the mouse Pc homolog at some time during the evolutionary process. Although these regions do not show significant sequence alignment when the relevant PcG proteins are compared, it does appear that the overall amino acid content has been maintained. Every organism we examined contains at least one component of PRC1 that has the characteristic high positive charge. An important test of the hypothesis that this charge distribution is central to compaction will involve testing the extent of compaction produced by the various PRC1 proteins from different species.

The role of PRC1 in compacting chromatin in vitro has been well established by our previous work, and the ability of *Drosophila* Psc to compact chromatin in vitro has been correlated with its ability to repress homeotic genes in vivo (King et al. 2005). More recent studies support the hypothesis that compaction is a key function of the PcG system and that PRC1 is a central player in this activity. A study of nucleosome turnover revealed slower rates of turnover on genes repressed by the PcG system than on genes bound by Trithorax-group protein, consistent with a more inaccessible state (Deal et al. 2010). A study that looked at compaction at the cytological level showed that the HOX clusters in mouse are compacted when repressed and that compaction is dependent on PRC1, but not on the

ubiquitylation function of PRC1 (Eskeland et al. 2010). This result separates the compaction and enzymatic functions of PRC1 and suggests that compaction has the primary role in the repression of HOX. The compaction observed in these studies in mice, measured cytologically, might be caused by the same PRC1 function that represses remodeling and creates compacted structures visible by electron microscopy. It is important to recognize that these two events have not yet been equated mechanistically.

The swapping of the compaction domain between Psc and Pc homologs has important ramifications for the interplay between histone H2A ubiquitylation and compaction. The ubiquitylation activity requires a Psc homolog and a RING homolog. There are at least two distinct PRC1 family complexes that contain both Psc and RING homologs in *Drosophila* and mammals. The originally defined PRC1 complex contains Psc, Pc, Pc, and RING homologs, whereas the dRAF complex (*Drosophila*) and BCOR complex (mammals) contain Psc and RING homologs but are not known to contain Pc homologs. Interestingly, the inclusion of the highly positively charged Psc in the dRAF complex is accompanied by the presence of the highly negatively charged Mtor and Ulp1, with the whole complex having a net charge of -17.3 . These latter complexes in the PRC1 family are proficient at ubiquitylation and have been proposed to be the central ubiquitylating complexes in *Drosophila* and mammals. The swap of the compaction domain between Psc and Pc means that in mammals, the compaction activity is found in only one subset of the PRC1 complexes, whereas in *Drosophila*, this domain is found in both PRC1 and dRAF. In mammals and in most organisms listed in our evolutionary comparison, there is the potential to separate these two functions between complexes. One can imagine that inclusion or omission of certain PcG functions could fine-tune the level of either compaction or ubiquitylation at the various genes regulated by the PRC1 family of complexes.

We have also mapped the domains involved in the association of the Zeste protein with PRC1 in *Drosophila* but were unable to uncover a clear functional role for that interaction. The Zeste protein is of significant interest in the biology of PRC1 because it binds stoichiometrically to the complex, both during reconstitution and when PRC1 is isolated from embryos, and it has a rich history of involvement in gene regulation, as determined by genetic studies (Pirrotta 1991). It therefore seems a strong candidate to have an important role in regulating PRC1 function. The types of role(s) that it might have are perplexing, because the genetic studies implicate it in both activation and repression, and it does not display classic PcG phenotypes and, in fact, displays some of the opposing Trithorax-group phenotypes.

Zeste interacts with Ph, a PRC1 subunit that is involved in bridging interactions either involving self-association or association with proteins such as Scm. Zeste is also able to self-associate. It is possible, then, that Zeste might modulate higher-order interactions between PcG complexes and targets. We tested whether Zeste might be involved in regulating compaction on arrays, possibly stimulated by changes in association, but did not observe any striking effects when

Zeste was either combined with Ph alone or mixed with a full PRC1 reconstituted complex. In both instances, Zeste increased the ability to repress remodeling, but the effects were of modest magnitude. Perhaps a simple possibility explains the interaction of Zeste with PRC1: Zeste has the ability to bind DNA with sequence specificity, and it might solely target the complex without affecting activity. Its intricate genetic functions, the involvement in *trans*-sensing and -acting phenomena like transvection, and its ability to interact with the Brahma complex, a Trithorax-group complex, indicate, on the other hand, that something more complex might be involved.

ACKNOWLEDGMENTS

We thank Brad Chapman for help with the phylogenetic analysis of PcG proteins and S.K. Bowman for critical reading of the manuscript. Funding for this work was provided by the National Institutes of Health (GM43901). J.M.A. was supported in part by a fellowship from Fundação para a Ciência e Tecnologia (SFRH/BD/11800/2003).

REFERENCES

- Abmayr SM, Yao T, Parmely T, Workman JL. 2001. *Preparation of nuclear and cytoplasmic extracts from mammalian cells*. Wiley, New York.
- Bernstein E, Duncan EM, Masui O, Gil J, Heard E, Allis CD. 2006. Mouse Polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol* **26**: 2560–2569.
- Biggin MD, Bickel S, Benson M, Pirrotta V, Tjian R. 1988. Zeste encodes a sequence-specific transcription factor that activates the Ultrabithorax promoter in vitro. *Cell* **53**: 713–722.
- Chen JD, Pirrotta V. 1993. Multimerization of the *Drosophila* zeste protein is required for efficient DNA binding. *EMBO J* **12**: 2075–2083.
- Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. 2002. *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**: 185–196.
- Deal RB, Henikoff JG, Henikoff S. 2010. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**: 1161–1164.
- Déjardin J, Cavalli G. 2004. Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. *EMBO J* **23**: 857–868.
- Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultschi AI, Wutz A, Bickmore WA. 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell* **38**: 452–464.
- Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**: 1870–1881.
- Francis NJ, Saurin AJ, Shao Z, Kingston RE. 2001. Reconstitution of a functional core Polycomb repressive complex. *Mol Cell* **8**: 545–556.
- Francis NJ, Kingston RE, Woodcock CL. 2004. Chromatin compaction by a Polycomb group protein complex. *Science* **306**: 1574–1577.
- Gearhart MD, Corcoran CM, Wamstad JA, Bardwell VJ. 2006. Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol Cell Biol* **26**: 6880–6889.
- Hur MW, Laney JD, Jeon SH, Ali J, Biggin MD. 2002. Zeste maintains repression of Ubx transgenes: Support for a new model of Polycomb repression. *Development* **129**: 1339–1343.
- Judd BH. 1995. Mutations of *zeste* that mediate transvection are recessive enhancers of position-effect variegation in *Drosophila melanogaster*. *Genetics* **141**: 245–253.
- Kal AJ, Mahmoudi T, Zak NB, Verrijzer CP. 2000. The *Drosophila* brahma complex is an essential coactivator for the trithorax group protein Zeste. *Genes Dev* **14**: 1058–1071.
- Karakuzu O, Wang DP, Cameron S. 2009. MIG-32 and SPAT-3A are PRC1 homologs that control neuronal migration in *Caenorhabditis elegans*. *Development* **136**: 943–953.
- King IF, Emmons RB, Francis NJ, Wild B, Müller J, Kingston RE, Wu CT. 2005. Analysis of a Polycomb group protein defines regions that link repressive activity on nucleosomal templates to in vivo function. *Mol Cell Biol* **25**: 6578–6591.
- Klymenko T, Papp B, Fischle W, Köcher T, Schelder M, Fritsch C, Wild B, Wilm M, Müller J. 2006. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev* **20**: 1110–1122.
- Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**: 2893–2905.
- Lagarou A, Mohd-Sarip A, Moshkin YM, Chalkley GE, Bezstarosti K, Demmers JA, Verrijzer CP. 2008. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev* **22**: 2799–2810.
- Lewis P. 1947. New mutants: Report of P. Lewis. *Drosophila Inform Serv* **21**: 69.
- Lewis EB. 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**: 565–570.
- Mulholland NM, King IF, Kingston RE. 2003. Regulation of Polycomb group complexes by the sequence-specific DNA binding proteins Zeste and GAGA. *Genes Dev* **17**: 2741–2746.
- Müller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. 2002. Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**: 197–208.
- Nakagawa T, Kajitani T, Togo S, Masuko N, Ohdan H, Hishikawa Y, Koji T, Matsuyama T, Ikura T, Muramatsu M, Ito T. 2008. Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes Dev* **22**: 37–49.
- Phelan ML, Sif S, Narlikar GJ, Kingston RE. 1999. Reconstitution of a core chromatin remodeling complex from SWI/SNF subunits. *Mol Cell* **3**: 247–253.
- Pirrotta V. 1991. The genetics and molecular biology of zeste in *Drosophila melanogaster*. *Adv Genet* **29**: 301–348.
- Ringrose L, Paro R. 2007. Polycomb/Trithorax response elements and epigenetic memory of cell identity. *Development* **134**: 223–232.
- Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY. 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**: 1311–1323.
- Satijn DP, Gunster MJ, van der Vlag J, Hamer KM, Schul W, Alkema MJ, Saurin AJ, Freemont PS, van Driel R, Otte AP. 1997. RING1 is associated with the Polycomb group protein complex and acts as a transcriptional repressor. *Mol Cell Biol* **17**: 4105–4113.
- Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P, Kingston RE. 2001. A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins. *Nature* **412**: 655–660.
- Scheuermann JC, de Ayala Alonso AG, Oktaba K, Ly-Hartig N, McGinty RK, Fraterman S, Wilm M, Muir TW, Müller J. 2010. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**: 243–247.
- Schoorlemmer J, Marcos-Gutiérrez C, Were F, Martínez R, García E, Satijn DP, Otte AP, Vidal M. 1997. Ring1A is a transcriptional repressor that interacts with the Polycomb-M33 protein and is expressed at rhombomere boundaries in the mouse hindbrain. *EMBO J* **16**: 5930–5942.
- Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE. 1999. Stabilization of chromatin structure by

- PRC1, a Polycomb complex. *Cell* **98**: 37–46.
- Sif S, Saurin AJ, Imbalzano AN, Kingston RE. 2001. Purification and characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes. *Genes Dev* **15**: 603–618.
- Simon JA, Kingston RE. 2009. Mechanisms of Polycomb gene silencing: Knowns and unknowns. *Nat Rev Mol Cell Biol* **10**: 697–708.
- Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, Koseki H, Brockdorff N, Fisher AG, Pombo A. 2007. Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**: 1428–1435.
- Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**: 750–756.

References

2011. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res* **39**: D214-219.
- Agianian B, Leonard K, Bonte E, Van der Zandt H, Becker PB, Tucker PA. 1999. The glutamine-rich domain of the Drosophila GAGA factor is necessary for amyloid fibre formation in vitro, but not for chromatin remodelling. *J Mol Biol* **285**: 527-544.
- Akam M. 1987. The molecular basis for metameric pattern in the Drosophila embryo. *Development* **101**: 1-22.
- Akasaka T, Kanno M, Balling R, Mieza MA, Taniguchi M, Koseki H. 1996. A role for mel-18, a Polycomb group-related vertebrate gene, during theanteriorposterior specification of the axial skeleton. *Development* **122**: 1513-1522.
- Akasaka T, van Lohuizen M, van der Lugt N, Mizutani-Koseki Y, Kanno M, Taniguchi M, Vidal M, Alkema M, Berns A, Koseki H. 2001. Mice doubly deficient for the Polycomb Group genes Mel18 and Bmi1 reveal synergy and requirement for maintenance but not initiation of Hox gene expression. *Development* **128**: 1587-1597.
- Alkema MJ, Bronk M, Verhoeven E, Otte A, van 't Veer LJ, Berns A, van Lohuizen M. 1997. Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian polycomb complex. *Genes Dev* **11**: 226-240.
- Armache KJ, Garlick JD, Canzio D, Narlikar GJ, Kingston RE. 2011. Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* **334**: 977-982.
- Beisel C, Paro R. 2011. Silencing chromatin: comparing modes and mechanisms. *Nat Rev Genet* **12**: 123-135.
- Bel S, Core N, Djabali M, Kieboom K, Van der Lugt N, Alkema MJ, Van Lohuizen M. 1998. Genetic interactions and dosage effects of Polycomb group genes in mice. *Development* **125**: 3543-3551.

- Bell O, Schwaiger M, Oakeley EJ, Lienert F, Beisel C, Stadler MB, Schubeler D. 2010. Accessibility of the *Drosophila* genome discriminates PcG repression, H4K16 acetylation and replication timing. *Nat Struct Mol Biol* **17**: 894-900.
- Biason-Lauber A, Konrad D, Meyer M, DeBeaufort C, Schoenle EJ. 2009. Ovaries and female phenotype in a girl with 46,XY karyotype and mutations in the CBX2 gene. *Am J Hum Genet* **84**: 658-663.
- Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK et al. 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**: 349-353.
- Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. 2006. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**: 1123-1136.
- Breiling A, Bonte E, Ferrari S, Becker PB, Paro R. 1999. The *Drosophila* polycomb protein interacts with nucleosomal core particles In vitro via its repression domain. *Mol Cell Biol* **19**: 8451-8460.
- Brown JL, Fritsch C, Mueller J, Kassis JA. 2003. The *Drosophila* pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. *Development* **130**: 285-294.
- Brown JL, Mucci D, Whiteley M, Dirksen ML, Kassis JA. 1998. The *Drosophila* Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol Cell* **1**: 1057-1064.
- Brunk BP, Martin EC, Adler PN. 1991. *Drosophila* genes Posterior Sex Combs and Suppressor two of zeste encode proteins with homology to the murine bmi-1 oncogene. *Nature* **353**: 351-353.
- Burgess RR. 2009. Protein precipitation techniques. *Methods Enzymol* **463**: 331-342.
- Cao R, Tsukada YI, Zhang Y. 2005. Role of Bmi-1 and Ring1A in H2A Ubiquitylation and Hox Gene Silencing. *Mol Cell* **20**: 845-854.

- Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**: 1422-1423.
- Core N, Bel S, Gaunt SJ, Aurrand-Lions M, Pearce J, Fisher A, Djabali M. 1997. Altered cellular proliferation and mesoderm patterning in Polycomb-M33-deficient mice. *Development* **124**: 721-729.
- Deal RB, Henikoff JG, Henikoff S. 2010. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**: 1161-1164.
- Dellino GI, Schwartz YB, Farkas G, McCabe D, Elgin SC, Pirrotta V. 2004. Polycomb silencing blocks transcription initiation. *Mol Cell* **13**: 887-893.
- Driever W, Nusslein-Volhard C. 1988a. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell* **54**: 95-104.
- . 1988b. A gradient of bicoid protein in Drosophila embryos. *Cell* **54**: 83-93.
- Duncan I. 1987. The bithorax complex. *Annu Rev Genet* **21**: 285-319.
- Eddy SR. 2009. A new generation of homology search tools based on probabilistic inference. *Genome Inform* **23**: 205-211.
- Emmons RB, Genetti H, Filandrinis S, Lokere J, Wu CT. 2009. Molecular genetic analysis of Suppressor 2 of zeste identifies key functional domains. *Genetics* **182**: 999-1013.
- Eskeland R, Leeb M, Grimes GR, Kress C, Boyle S, Sproul D, Gilbert N, Fan Y, Skoultschi AI, Wutz A et al. 2010. Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell* **38**: 452-464.

- Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. 2003. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev* **17**: 1870-1881.
- Forzati F, Federico A, Pallante P, Abbate A, Esposito F, Malapelle U, Sepe R, Palma G, Troncone G, Scarfo M et al. 2012. CBX7 is a tumor suppressor in mice and humans. *J Clin Invest*.
- Francis NJ, Follmer NE, Simon MD, Aghia G, Butler JD. 2009. Polycomb proteins remain bound to chromatin and DNA during DNA replication in vitro. *Cell* **137**: 110-122.
- Francis NJ, Kingston RE, Woodcock CL. 2004. Chromatin compaction by a polycomb group protein complex. *Science* **306**: 1574-1577.
- Francis NJ, Saurin AJ, Shao Z, Kingston RE. 2001. Reconstitution of a functional core polycomb repressive complex. *Mol Cell* **8**: 545-556.
- Franke A, Messmer S, Paro R. 1995. Mapping functional domains of the polycomb protein of *Drosophila melanogaster*. *Chromosome Res* **3**: 351-360.
- Garcia E, Marcos-Gutierrez C, del Mar Lorente M, Moreno JC, Vidal M. 1999. RYBP, a new repressor protein that interacts with components of the mammalian Polycomb complex, and with the transcription factor YY1. *EMBO J* **18**: 3404-3418.
- Garza AS, Ahmad N, Kumar R. 2009. Role of intrinsically disordered protein regions/domains in transcriptional regulation. *Life Sci* **84**: 189-193.
- Gaytan de Ayala Alonso A, Gutierrez L, Fritsch C, Papp B, Beuchle D, Muller J. 2007. A genetic screen identifies novel polycomb group genes in *Drosophila*. *Genetics* **176**: 2099-2108.
- Gunasekaran K, Tsai CJ, Kumar S, Zanuy D, Nussinov R. 2003. Extended disordered proteins: targeting function with less scaffold. *Trends Biochem Sci* **28**: 81-85.

- Gunster MJ, Satijn DP, Hamer KM, den Blaauwen JL, de Bruijn D, Alkema MJ, van Lohuizen M, van Driel R, Otte AP. 1997. Identification and characterization of interactions between the vertebrate polycomb-group protein BMI1 and human homologs of polyhomeotic. *Mol Cell Biol* **17**: 2326-2335.
- Gutierrez L, Oktaba K, Scheuermann JC, Gambetta MC, Ly-Hartig N, Muller J. 2012. The role of the histone H2A ubiquitinase Scc in Polycomb repression. *Development* **139**: 117-127.
- Hansen JC, Lu X, Ross ED, Woody RW. 2006. Intrinsic protein disorder, amino acid composition, and histone terminal domains. *J Biol Chem* **281**: 1853-1856.
- Hansen KH, Bracken AP, Pasini D, Dietrich N, Gehani SS, Monrad A, Rappsilber J, Lerdrup M, Helin K. 2008. A model for transmission of the H3K27me3 epigenetic mark. *Nat Cell Biol* **10**: 1291-1300.
- Hashimoto N, Brock HW, Nomura M, Kyba M, Hodgson J, Fujita Y, Takihara Y, Shimada K, Higashinakagawa T. 1998. RAE28, BMI1, and M33 are members of heterogeneous multimeric mammalian Polycomb group complexes. *Biochem Biophys Res Commun* **245**: 356-365.
- Haupt Y, Alexander WS, Barri G, Klinken SP, Adams JM. 1991. Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice. *Cell* **65**: 753-763.
- He A, Shen X, Ma Q, Cao J, von Gise A, Zhou P, Wang G, Marquez VE, Orkin SH, Pu WT. 2012. PRC2 directly methylates GATA4 and represses its transcriptional activity. *Genes Dev* **26**: 37-42.
- Hemenway CS, Halligan BW, Levy LS. 1998. The Bmi-1 oncoprotein interacts with dinG and MPh2: the role of RING finger domains. *Oncogene* **16**: 2541-2547.
- Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L et al. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**: D211-215.

- Jürgens G. 1985. A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature* **316**: 153-155.
- Kanno M, Hasegawa M, Ishida A, Isono K, Taniguchi M. 1995. mel-18, a Polycomb group-related mammalian gene, encodes a transcriptional negative regulator with tumor suppressive activity. *EMBO J* **14**: 5672-5678.
- Kathuria SV, Guo L, Graceffa R, Barrea R, Nobrega RP, Matthews CR, Irving TC, Bilsel O. 2011. Minireview: structural insights into early folding events using continuous-flow time-resolved small-angle X-ray scattering. *Biopolymers* **95**: 550-558.
- Katoh-Fukui Y, Miyabayashi K, Komatsu T, Owaki A, Baba T, Shima Y, Kidokoro T, Kanai Y, Schedl A, Wilhelm D et al. 2011. Cbx2, a Polycomb Group Gene, Is Required for Sry Gene Expression in Mice. *Endocrinology*.
- Katoh-Fukui Y, Tsuchiya R, Shiroishi T, Nakahara Y, Hashimoto N, Noguchi K, Higashinakagawa T. 1998. Male-to-female sex reversal in M33 mutant mice. *Nature* **393**: 688-692.
- Kaufman TC, Seeger MA, Olsen G. 1990. Molecular and genetic organization of the antennapedia gene complex of *Drosophila melanogaster*. *Adv Genet* **27**: 309-362.
- Kelly SM, Price NC. 2000. The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci* **1**: 349-384.
- Keppel TR, Howard BA, Weis DD. 2011. Mapping unstructured regions and synergistic folding in intrinsically disordered proteins with amide H/D exchange mass spectrometry. *Biochemistry* **50**: 8722-8732.
- King IF, Emmons RB, Francis NJ, Wild B, Muller J, Kingston RE, Wu CT. 2005. Analysis of a polycomb group protein defines regions that link repressive activity on nucleosomal templates to in vivo function. *Mol Cell Biol* **25**: 6578-6591.
- King IF, Francis NJ, Kingston RE. 2002. Native and recombinant polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro. *Mol Cell Biol* **22**: 7919-7928.

- Klymenko T, Papp B, Fischle W, Kocher T, Schelder M, Fritsch C, Wild B, Wilm M, Muller J. 2006. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev* **20**: 1110-1122.
- Kurowski MA, Bujnicki JM. 2003. GeneSilico protein structure prediction meta-server. *Nucleic Acids Res* **31**: 3305-3307.
- Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. 2002. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**: 2893-2905.
- Kwon H, Imbalzano AN, Khavari PA, Kingston RE, Green MR. 1994. Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* **370**: 477-481.
- Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**: 116-120.
- Lagarou A, Mohd-Sarip A, Moshkin YM, Chalkley GE, Bezstarosti K, Demmers JA, Verrijzer CP. 2008. dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev* **22**: 2799-2810.
- Lavigne M, Francis NJ, King IF, Kingston RE. 2004. Propagation of silencing; recruitment and repression of naive chromatin in trans by polycomb repressed chromatin. *Mol Cell* **13**: 415-425.
- Levine SS, Weiss A, Erdjument-Bromage H, Shao Z, Tempst P, Kingston RE. 2002. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol* **22**: 6070-6078.
- Lewis EB. 1947. New mutants: Report of P. Lewis. *Drosophila Inform Serv* **21**: 69.
- . 1978. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**: 565-570.

- Lippens G, Sillen A, Smet C, Wieruszeski JM, Leroy A, Buee L, Landrieu I. 2006. Studying the natively unfolded neuronal Tau protein by solution NMR spectroscopy. *Protein Pept Lett* **13**: 235-246.
- Lo SM, Ahuja NK, Francis NJ. 2009. Polycomb group protein Suppressor 2 of zeste is a functional homolog of Posterior Sex Combs. *Mol Cell Biol* **29**: 515-525.
- Lo SM, Francis NJ. 2010. Inhibition of chromatin remodeling by polycomb group protein posterior sex combs is mechanistically distinct from nucleosome binding. *Biochemistry* **49**: 9438-9448.
- Lund AH, van Lohuizen M. 2004. Polycomb complexes and silencing mechanisms. *Curr Opin Cell Biol* **16**: 239-246.
- Ma J, Ptashne M. 1987. A new class of yeast transcriptional activators. *Cell* **51**: 113-119.
- Margueron R, Li G, Sarma K, Blais A, Zavadil J, Woodcock CL, Dynlacht BD, Reinberg D. 2008. Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol Cell* **32**: 503-518.
- Meadow ND, Fox DK, Roseman S. 1990. The bacterial phosphoenolpyruvate: glucose phosphotransferase system. *Annu Rev Biochem* **59**: 497-542.
- Messmer S, Franke A, Paro R. 1992. Analysis of the functional role of the Polycomb chromo domain in *Drosophila melanogaster*. *Genes Dev* **6**: 1241-1254.
- Min J, Zhang Y, Xu RM. 2003. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev* **17**: 1823-1828.
- Mohd-Sarip A, Cleard F, Mishra RK, Karch F, Verrijzer CP. 2005. Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. *Genes Dev* **19**: 1755-1760.
- Morey L, Pascual G, Cozzuto L, Roma G, Wutz A, Benitah SA, Di Croce L. 2012. Nonoverlapping functions of the polycomb group cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **10**: 47-62.

- Mousavi K, Zare H, Wang AH, Sartorelli V. 2011. Polycomb Protein Ezh1 Promotes RNA Polymerase II Elongation. *Mol Cell*.
- Muller J. 1995. Transcriptional silencing by the Polycomb protein in Drosophila embryos. *EMBO J* **14**: 1209-1220.
- Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. 2002. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**: 197-208.
- Muller J, Kassis JA. 2006. Polycomb response elements and targeting of Polycomb group proteins in Drosophila. *Curr Opin Genet Dev* **16**: 476-484.
- Muller J, Verrijzer P. 2009. Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr Opin Genet Dev* **19**: 150-158.
- Neira JL, Roman-Trufero M, Contreras LM, Prieto J, Singh G, Barrera FN, Renart ML, Vidal M. 2009. The transcriptional repressor RYBP is a natively unfolded protein which folds upon binding to DNA. *Biochemistry* **48**: 1348-1360.
- O'Loughlen A, Munoz-Cabello AM, Gaspar-Maia A, Wu HA, Banito A, Kunowska N, Racek T, Pemberton HN, Beolchi P, Laval F et al. 2012. MicroRNA Regulation of Cbx7 Mediates a Switch of Polycomb Orthologs during ESC Differentiation. *Cell Stem Cell* **10**: 33-46.
- Paro R, Hogness DS. 1991. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc Natl Acad Sci U S A* **88**: 263-267.
- Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, Bak M, Tommerup N, Rappsilber J, Helin K. 2010. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**: 306-310.
- Pattatucci AM, Kaufman TC. 1991. The homeotic gene Sex combs reduced of Drosophila melanogaster is differentially regulated in the embryonic and imaginal stages of development. *Genetics* **129**: 443-461.

- Pattatucci AM, Otteson DC, Kaufman TC. 1991. A functional and structural analysis of the Sex combs reduced locus of *Drosophila melanogaster*. *Genetics* **129**: 423-441.
- Pearce JJ, Singh PB, Gaunt SJ. 1992. The mouse has a Polycomb-like chromobox gene. *Development* **114**: 921-929.
- Peng JC, Valouev A, Swigut T, Zhang J, Zhao Y, Sidow A, Wysocka J. 2009. Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**: 1290-1302.
- Ren X, Vincenz C, Kerppola TK. 2008. Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. *Mol Cell Biol* **28**: 2884-2895.
- Sandhu KS. 2009. Intrinsic disorder explains diverse nuclear roles of chromatin remodeling proteins. *J Mol Recognit* **22**: 1-8.
- Sanson B. 2001. Generating patterns from fields of cells. Examples from *Drosophila* segmentation. *EMBO Rep* **2**: 1083-1088.
- Sarma K, Margueron R, Ivanov A, Pirrotta V, Reinberg D. 2008. Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol Cell Biol* **28**: 2718-2731.
- Satijn DP, Hamer KM, den Blaauwen J, Otte AP. 2001. The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in *Xenopus* embryos. *Mol Cell Biol* **21**: 1360-1369.
- Satijn DP, Otte AP. 1999. RING1 interacts with multiple Polycomb-group proteins and displays tumorigenic activity. *Mol Cell Biol* **19**: 57-68.
- Sato T, Russell MA, Denell RE. 1983. Homoeosis in *Drosophila*: a new enhancer of polycomb and related homoeotic mutations. *Genetics* **105**: 357-370.

- Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P, Kingston RE. 2001. A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature* **412**: 655-660.
- Scheuermann JC, de Ayala Alonso AG, Oktaba K, Ly-Hartig N, McGinty RK, Fraterman S, Wilm M, Muir TW, Muller J. 2010. Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. *Nature* **465**: 243-247.
- Schoorlemmer J, Marcos-Gutierrez C, Were F, Martinez R, Garcia E, Satijn DP, Otte AP, Vidal M. 1997. Ring1A is a transcriptional repressor that interacts with the Polycomb-M33 protein and is expressed at rhombomere boundaries in the mouse hindbrain. *Embo J* **16**: 5930-5942.
- Schwartz YB, Kahn TG, Nix DA, Li XY, Bourgon R, Biggin M, Pirrotta V. 2006. Genome-wide analysis of Polycomb targets in Drosophila melanogaster. *Nat Genet* **38**: 700-705.
- Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE. 1999. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* **98**: 37-46.
- Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, Yuan GC, Lee Y, Orkin SH. 2009. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* **139**: 1303-1314.
- Shen X, Liu Y, Hsu YJ, Fujiwara Y, Kim J, Mao X, Yuan GC, Orkin SH. 2008. EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol Cell* **32**: 491-502.
- Shi Y, Seto E, Chang LS, Shenk T. 1991. Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* **67**: 377-388.
- Shoemaker BA, Portman JJ, Wolynes PG. 2000. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* **97**: 8868-8873.

- Sif S, Saurin AJ, Imbalzano AN, Kingston RE. 2001. Purification and characterization of mSin3A-containing Brg1 and hBrm chromatin remodeling complexes. *Genes Dev* **15**: 603-618.
- Sif S, Stukenberg PT, Kirschner MW, Kingston RE. 1998. Mitotic inactivation of a human SWI/SNF chromatin remodeling complex. *Genes Dev* **12**: 2842-2851.
- Sigler PB. 1988. Transcriptional activation. Acid blobs and negative noodles. *Nature* **333**: 210-212.
- Simon J, Chiang A, Bender W. 1992. Ten different Polycomb group genes are required for spatial control of the abdA and AbdB homeotic products. *Development* **114**: 493-505.
- Simon J, Chiang A, Bender W, Shimell MJ, O'Connor M. 1993. Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. *Dev Biol* **158**: 131-144.
- Struhl G, Akam M. 1985. Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. *EMBO J* **4**: 3259-3264.
- Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* **41**: 207-234.
- Suganuma T, Workman JL. 2011. Signals and combinatorial functions of histone modifications. *Annu Rev Biochem* **80**: 473-499.
- Utlei RT, Ikeda K, Grant PA, Cote J, Steger DJ, Eberharter A, John S, Workman JL. 1998. Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature* **394**: 498-502.
- Uversky VN, Dunker AK. 2010. Understanding protein non-folding. *Biochim Biophys Acta* **1804**: 1231-1264.

- van der Lugt NM, Domen J, Linders K, van Roon M, Robanus-Maandag E, te Riele H, van der Valk M, Deschamps J, Sofroniew M, van Lohuizen M et al. 1994. Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev* **8**: 757-769.
- van Lohuizen M, Frasch M, Wientjens E, Berns A. 1991a. Sequence similarity between the mammalian bmi-1 proto-oncogene and the Drosophila regulatory genes Psc and Su(z)2. *Nature* **353**: 353-355.
- van Lohuizen M, Verbeek S, Scheijen B, Wientjens E, van der Gulden H, Berns A. 1991b. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell* **65**: 737-752.
- Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y. 2004. Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**: 873-878.
- Wang R, Ilango U, Robinson AK, Schirf V, Schwarz PM, Lafer EM, Demeler B, Hinck AP, Kim CA. 2008. Structural transitions of the RING1B C-terminal region upon binding the polycomb cbox domain. *Biochemistry* **47**: 8007-8015.
- Woo CJ, Kharchenko PV, Daheron L, Park PJ, Kingston RE. 2010. A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* **140**: 99-110.
- Wu CT, Howe M. 1995. A genetic analysis of the Suppressor 2 of zeste complex of *Drosophila melanogaster*. *Genetics* **140**: 139-181.